DOCUMENT RESUME

ED 479 162 HE 036 069

AUTHOR	Metcalf, Kim K.; West, Stephen D.; Legan, Natalie; Paul, Kelli; Boone, William J.
TITLE	Evaluation of the Cleveland Scholarship and Tutoring Program, 1998-2001. Summary Report [and] Technical Report.
INSTITUTION	Indiana Univ., Bloomington. Indiana Center for Evaluation.
PUB DATE	2002-10-00
NOTE	216p.; Executive summary is included. For the summary report through 2000, see UD 034 881.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	EDRS Price MF01/PC09 Plus Postage.
DESCRIPTORS	Annual Reports; *Educational Vouchers; Elementary Secondary Education; Private Schools; Public Schools; Racial Differences; *Scholarship; *School Choice; Tutorial Programs; Urban Schools; *Urban Youth

ABSTRACT

This is the second annual report of a study examining the operation and impact of the Cleveland Scholarship and Tutoring Program (CSTP). The longitudinal project is investigating a wide range of issues associated with the operation and impact of the publicly funded voucher program. For the current project year, the evaluation focused on the characteristics of participating students, teachers, and classrooms at the private schools students attend and the public schools, and the academic achievement or program participants. This report is based on data collection and analysis from autumn 1998 through April 2001 and focuses on a cohort that began schooling as first graders in the 1998-1999 academic year. These children include scholarship students (voucher recipients), applicant/nonrecipients, and nonapplicants. Of secondary interest are scholarship winners/nonusers and former scholarship students who returned to the public schools. A greater proportion of scholarship students are Caucasian and fewer are African American than public school students, but the proportion of Hispanic and multiracial scholarship students is nearly twice that of public school students. The vast majority of teachers at public and private schools are fully certified, although the proportion is slightly greater in the public schools. Class sizes were larger in private schools, and larger classes were associated with higher academic achievement. There were no consistent, significant differences in achievement between scholarship and public school students by the end of third grade, and this finding held true across all of the available achievement measures. Students who leave the program and enroll in the public schools achieve at the lowest level of any other group. The findings reinforce that policies and procedures for disseminating scholarships have important implications for the program and participating families. Three appendixes contain information about analytic techniques. (Contains 62 tables and 32 figures.) (SLD)

Indiana Center for **Evaluation**

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

Evaluation of the Cleveland Scholarship and Tutoring Program 1998-2001

EXECUTIVE SUMMARY

October, 2002

This is the second annual report of a study focused on examining the operation and impact of the Cleveland Scholarship and Tutoring Program (CSTP). The longitudinal project is investigating a wide range of issues associated with the operation and impact of the publicly-funded voucher program. However, in the current project year, the study was guided primarily by three questions:

1. What are the characteristics of students who participate in the CSTP, and how do they compare with students who do not participate?
2. What are the characteristics of the classrooms and teachers with whom scholarship students work in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?
3. What is the impact of participation in the program on student academic achievement?

The current report presents the results of data collection and analysis over the period from autumn, 1998 through April, 2001. The project has been focused on a cohort of children who began their public or private schooling as first graders during the 1998-99 academic year and who were completing third grade in spring, 2001. These children represent three primary and two secondary groups of interest, and they provide multiple groups with whom scholarship (voucher) students may be compared. Primary groups include: (1) *scholarship students* who were awarded and use a voucher to attend private school; (2) *applicant/non-recipients*, who are public school students whose families applied for a scholarship but who were not selected to receive one; and (3) *non-applicants*, who are public school students whose families have never applied for a scholarship. Secondary groups include: (4) *scholarship winner/non-users*, who represent a small group of public school students whose families applied for and were awarded a voucher, but whose families elected not to use the scholarship; and (5) *former scholarship students*, who were awarded, accepted, and used a voucher for private school enrollment for one or more years, but subsequently withdrew from the program to return to public schools.

What follows are major findings drawn from the more detailed technical and summary reports. It should be noted that, unless otherwise indicated, findings are based on the longitudinal cohort of students who are now in third grade.

STUDENT CHARACTERISTICS

§ A greater proportion of scholarship students are Caucasian and fewer are African American than public school students. However, the proportion of Hispanic and multiracial scholarship students is nearly twice that of public school students.

- § Third grade scholarship students, in general, are less likely to qualify for free lunch than public school students.
- § Students in both scholarship and public school groups are nearly equally male and female and most have attended the same school from first through third grade.
- § Most students in the cohort who entered the program after first grade had unsuccessfully applied for a scholarship in prior year, and the proportion of prior applicants was roughly the same for students who had been enrolled in private and public schools.
- § Across grades, students who accept their scholarship prior to the beginning of the academic year are nearly identical to their public school classmates in race/ethnicity and income. However, a majority of students who accept their scholarship after the beginning of the school year consists of students who were already enrolled in private schools, are of higher income, and are less likely to be African American.

CLASSROOM AND TEACHER CHARACTERISTICS

- § The vast majority of teachers in both private and public schools are fully certified, although the proportion is slightly greater in public schools.
- § Private school teachers had more experience and had been in their current schools longer than public school teachers.
- § Public school teachers were more likely to have completed some graduate coursework, but a majority of teachers in both groups had done so.
- § Class sizes were larger in private schools, and class size was *positively* related to student achievement (i.e., larger classes were associated with higher achievement)
- § Minority students in both public and private schools tended to be enrolled in smaller classes and with more experienced teachers than non-minority students, but non-minority students were more likely to work with fully certified teachers.

STUDENT ACHIEVEMENT

- § There are no consistent, significant differences in achievement between scholarship and public school students by the end of third grade. This finding holds across all of the available achievement measures (reading, language arts, mathematics, social studies, science, and total battery).
- § Students who leave the program and enroll in public schools are achieving at lower levels than any other group in the study. Further, this pattern of comparatively low achievement generally continues for these students in public schools.

CONCLUSIONS

Findings from the most recent project period are not surprising. However, they do reinforce that policies and procedures for disseminating the scholarships have important implications for the program and the families who participate. Remaining unanswered are questions related to the long-term impact of the program on students, in academic and non-academic areas, and the context in which parents make choices about their children's education. Further, and notably, the security of the program that results from the recent Supreme Court decision likely will influence the program in important ways.

The longitudinal study of the voucher program in Cleveland represents the longest running and only ongoing research of any publicly-funded voucher program in the United States. As such, it uniquely offers the opportunity to provide answers to many of these questions.

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE SECTION 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

EVALUATION OF THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM

SUMMARY REPORT
1998 - 2001

Kim K. Metcalf, Ph.D.
Director

Stephen D. West, M.A.
Natalie A. Legan, M.S.
Kelli M. Paul, M.S

William J. Boone, Ph.D.

October, 2002

Indiana Center
for
Evaluation

509 East Third Street
Bloomington, Indiana 47401
http://www.indiana.edu/~iuice

Kim K. Metcalf, Ph.D.
Director

812-855-4438
800-511-6575
812-856-5890
iuice@indiana.edu

1998-2001 Summary Report

Evaluation of the Cleveland Scholarship and Tutoring Program (CSTP) was initiated by the Indiana Center for Evaluation in the spring of 1997 near the conclusion of the first operating year of the program. From the beginning, the evaluation has examined the impact of the CSTP on the students, families, and schools that participate, and the impact on the public schools from which the scholarship students are drawn. In November 1998, the evaluation began following a cohort of first grade scholarship and public school children, with the intention of collecting longitudinal data on the impact of the program.

The present report details the findings of data drawn from the cohort from the time the students began first grade in the fall of 1998 through the end of third grade, 2001. The sample under study includes three primary groups of students for whom sufficient data were available: (1) *scholarship students*, who are further classified based on the year during which they entered the program (i.e., kindergarteners in 1998-1999, first graders in 1999-2000, second graders in 2000-2001, or third graders in 2001-2002); (2) *applicant/non-recipients*, who are public school students whose families applied for a scholarship at some point between kindergarten,1998-1999, and third grade, 2001-2002, but who were not selected to receive a scholarship through the random lottery or because they did not meet (i.e., exceeded) the income requirement; and (3) *non-applicants*, who are public school students whose families have never applied for a scholarship at any point between kindergarten,1998-1999, and third grade, 2001-2002. As a result, the current project includes multiple comparison groups by which scholarship students' progress is weighed.

Two additional groups, comprised of relatively fewer students than the three primary groups described above, have been identified and subjected to analysis when appropriate. These groups are: (1) *scholarship winner/non-users*, who represent a small group public school students whose families applied for and were awarded a scholarship at some point between kindergarten and third grade, but whose families elected not to use the scholarship and remained in public schools; and (2) *former scholarship students,* who were awarded, accepted, and used a scholarship at some point between kindergarten and third grade, but subsequently withdrew from the program after one, two, or three years of participation to return to public schools.

During the period covered by the current report – which represents only a portion of the ongoing multiyear, longitudinal study of the voucher program in Cleveland – data collection and analysis have been guided by three questions associated with the characteristics of students, the characteristics of classrooms and teachers with whom they worked, and the impact of the CSTP on students' academic achievement.

Specifically, these questions are:

1. What are the characteristics of students who participate in the CSTP and how do they compare with students who do not participate?

2. What are the characteristics of the classrooms and teachers experienced by students who participate in the program, and how do they compare with those experienced by public school students?

3. What is the impact of participation in the program on students' academic achievement, and other relevant school related variables (e.g., attendance, behavior)?

During the period from 1998-2001, the evaluation emphasized the identification of complete and representative groups of students who constitute the longitudinal sample and on collecting initial data on questions 1, 2, and 3 above. A comprehensive technical report has been prepared in which the statistical and methodological techniques of the project are explained in detail. The present report attempts to summarize the basic evaluation activities in ways that avoid technical language and that focus on the findings and their interpretation. Individuals who would like more information on the details of

the study, including the specific measures used in collecting data and their operational definitions, are encouraged to obtain the longer technical report.

The remainder of this summary provides a brief overview of the findings from the evaluation of the Cleveland Scholarship and Tutoring Program conducted from fall, 1998 through spring, 2001, as well as a discussion of tentative conclusions that can be drawn from these findings. The summary is organized around the three evaluation questions noted above and for which data were collected and analyzed. Findings are frequently presented graphically rather than in tables and supplemented with narrative explanations for clarity. Several points should be made, however, prior to engaging in this discussion.

1. Unless otherwise noted, the results are based on a cohort of students who entered public or private school as first graders during the 1998-99 academic year. Data on these students were collected most recently in the spring of 2001, when cohort students were completing third grade. To the extent that this cohort of students is representative of the broader population of scholarship and public school students, which it generally appears to be, the findings can appropriately be generalized to the broader program. However, because the characteristics or experiences of this cohort may be unique in some ways from those of other students, generalizations must be made cautiously.

2. Any examination of the impacts of an educational program is a complex and difficult endeavor. This complexity is increased when the program is relatively large, encompasses a large number of schools, families, and students, and must be examined over time. Any attempt to draw broad or highly general conclusions across a range of findings requires a degree of simplification that cannot convey important nuances in the program or in the results.

3. The findings summarized below are only the most recent from an ongoing study of the Scholarship and Tutoring Program in Cleveland. As such, they represent only a snapshot of the program and the cohort of students as they were in May of 2001. Definitive, summative, or final conclusions about the program, schools, or students cannot and should not be drawn. Collectively, these conditions require that caution be exercised as findings are synthesized and conclusions drawn.

In addition, other researchers of publicly-funded vouchers, including our own earlier work, have produced differing findings and have reached vastly different conclusions.[1] We have speculated in other arenas that these divergent findings may be the result of differing methodological approaches to the study of voucher programs and to the

[1] All prior reports, from both phases one and two of the CSTP evaluation, are available through the Ohio Department of Education or through the Indiana Center for Evaluation (www.indiana.edu/~iuice/).

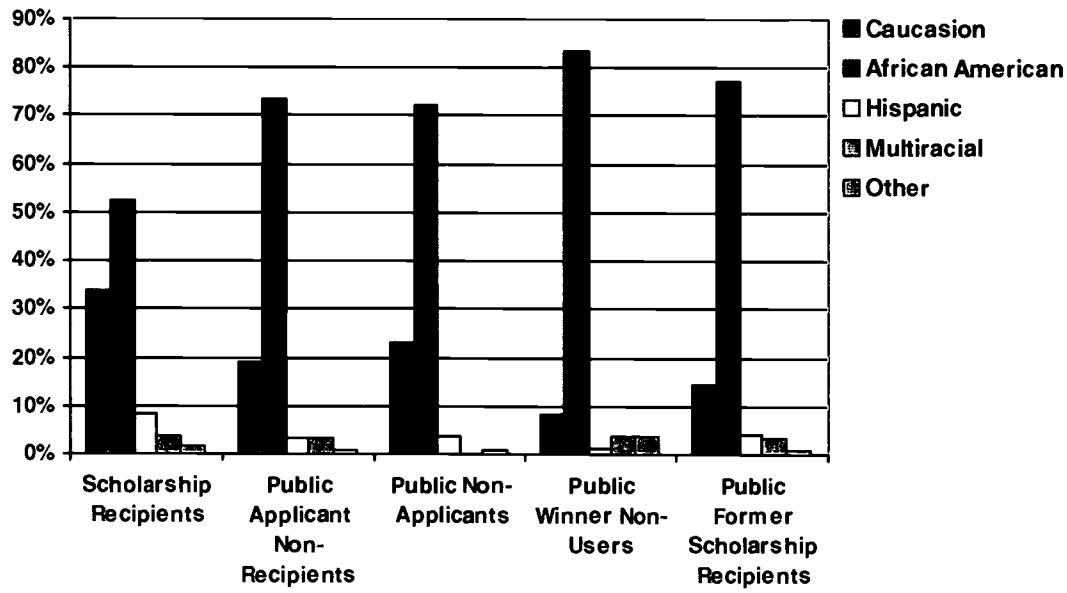
cumulative effects of education (and thus of voucher programs) that studies conducted over only three or four years may not identify (see Metcalf et al., 1999; Metcalf & Beghetto, 2001; Metcalf & Legan, 2001). The present study addresses these issues in ways that, we believe, may prove useful in reconciling these seemingly disparate results.

Evaluation Question One

What are the characteristics of students who participate in the Cleveland Scholarship and Tutoring Program, and how do they compare with students who do not participate?

In the earlier phase of our work in Cleveland, we found that scholarship students were somewhat more likely to be Caucasian and of much lower income than their public school counterparts. The current results offer a somewhat different picture. As in our earlier work, a greater proportion of the students in the third grade cohort of scholarship students are Caucasian and fewer are African-American in comparison to the proportions in public schools. However, the proportion of Hispanic and multiracial students in the third grade scholarship cohort is roughly twice that of the public schools. Thus, while the cohort is more Caucasian than the public schools, it is also slightly more diverse. Figure 1 depicts minority status across groups

Figure 1. 2000-2001 Minority Status by Category and Scholarship Group



Our current findings on family income of scholarship students also contrast with our earlier findings, which found no difference between scholarship students and the public school comparison groups. Data collected in the spring of 2001 indicate that the scholarship cohort is, in general, less likely to qualify for free lunch than public school students. Moreover, students who were awarded but choose not to use a scholarship are of substantially lower income than even their public school classmates. However, current findings regarding student sex and mobility were similar to our previous findings. That is, scholarship and public school students are nearly equally male and female, and the vast majority of students in both groups have been enrolled in the same school from first grade through third grade (i.e., relatively low student mobility across all groups).

More important, a clear pattern emerges in the characteristics of students who are awarded and accept their scholarship prior to the beginning of the academic year and those who do so after the beginning of school. Students who accept their scholarship prior to the beginning of school are nearly identical to their public school classmates in family income and race, and a majority is from public schools. However, a majority of students who accept their scholarship after the beginning of the school year consist of former private school students (likely already enrolled in the private school), are of higher income, and are less likely to be African-American.

On the basis of these findings, it seems that the random lottery, held by the CSTP office each spring, generally serves to award scholarships to students who have been in public schools and who are very similar, demographically, to their peers in Cleveland public schools. However, as the beginning of the academic year approaches and then passes, some students elect not to use scholarships that they were awarded. These scholarships are then offered to other applicants to the program (who may not have been selected in the initial lottery). Students who, at this later point, are awarded and choose to use a scholarship tend to be Caucasian, from families with higher incomes than earlier awardees, and have been enrolled in private school previously.

Additional analyses, which were possible for the first time in the current phase of the evaluation, revealed that students who enter the program after first grade tend to come from families who have applied for a voucher in prior years. There is also a slight tendency for them to have been enrolled in a private school before entering the CSTP.

However, and surprisingly, the proportion of new scholarship students whose families have previously applied for a scholarship is nearly identical for former public and former private school students. It appears that many families become interested in obtaining a scholarship early in their children's schooling, and they persist in their efforts to do so over multiple years.

Finally, the variables of student gender, race/ethnicity, income, and mobility, both individually and collectively, had only very weak relationships to the achievement of students in the cohort. Higher overall academic performance was generally associated with being Caucasian, of higher estimated family income, less mobile, and female. This trend was true both for scholarship and for public school students.

Evaluation Question Two

What are the characteristics of the classrooms and teachers with whom scholarship students work in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?

As in previous years, we found both similarities and differences between public and private school classrooms and teachers. The vast majority of teachers in both types of school were fully certified, although the proportion was greater in public schools than in private. Private school teachers had more experience and had been in their current schools longer than public school teachers. Public school teachers were more likely to have completed some graduate coursework, but most teachers in both types of school had done so. Class sizes were slightly larger in private schools, but both public and private school classrooms contained a greater number of students than many would recommend (e.g., the Tennessee STAR report). Of these variables, only class size was related to student achievement and this relationship, surprisingly, was positive (i.e., larger class sizes were associated with higher academic achievement).

In addition, an interesting trend was found in the classrooms experienced by minority and non-minority students in the cohort we have studied. Minority students tend to be enrolled in smaller classes and with teachers who have more experience, regardless of whether they attend private or public schools. However, non-minority students are more

likely to work with teachers who are fully certified. These findings, while significant, may be a result of any number of factors, such as unique characteristics of the third grade classrooms, unique characteristics of schools in districts with different proportions of minority and non-minority students, differential patterns of school choice between minority and non-minority families, or schools may attempt to secure more experienced teachers and smaller classes for minority students. Future years of the study will allow us to identify patterns that may exist.

Evaluation Question Three

What is the impact of participation in the Cleveland Scholarship and Tutoring Program on student academic achievement?

The most recent results do not reveal any significant impacts of participation in the CSTP on student achievement. From the end of first grade, when the large initial differences between public school students and scholarship students no longer existed, students in all of the groups we have studied demonstrated significant gains each year. And, across groups, the general extent of achievement growth was nearly equal through second and third grade. Although it is not statistically significant in the data available to date, there is some evidence of a pattern of slightly greater annual achievement growth among students who have used a scholarship continuously since kindergarten. If this pattern continues, the achievement of this group of students may become noticeably, and meaningfully, higher than that of public school students. However, data over three to five additional years will be necessary to confirm or discount such a pattern.

As we have found in previous years, students who choose to leave the CSTP and enroll in public schools are achieving at lower levels than any other group of students in our study. This trend is found regardless of the grade at which students discontinue the program. The data available on these students over multiple years now also indicate that this pattern of comparatively low achievement continues for these students in their public schools. A notable exception to this trend is the group of students who left the program after kindergarten and have attended public school consistently since that time. These students were achieving at roughly equal levels to other scholarship and public school students at the beginning of first grade, and continue to do so. For other students who

discontinue in the CSTP, future data collection will attempt to identify potential explanations either for their persistent lower achievement, or for improvement in their achievement after attending public schools for multiple years.

Figures 2, 3, 4, and 5 display the patterns of achievement for scholarship recipients, applicant non-recipients, non-applicants, and scholarship winner non-users. For these analyses, the scholarship recipient group has been further divided into four-year scholarship recipients (who have participated in the program continuously since they entered as kindergarteners in 1998-1999) and three-year scholarship recipients (who have participated in the program continuously since they entered as first graders in 1999-2000).

Figure 2. Reading Achievement: 1998 to 2001

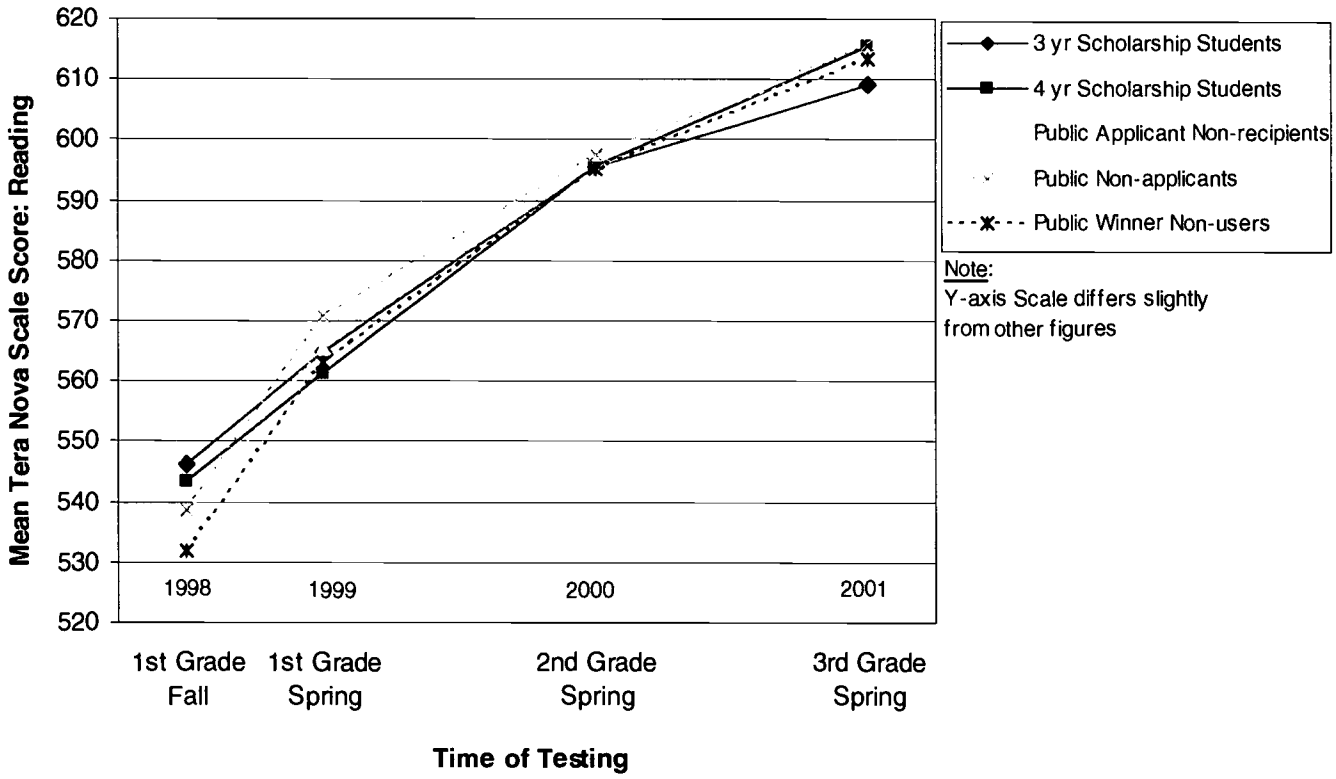


Figure 3. Language Achievement: 1998 to 2001

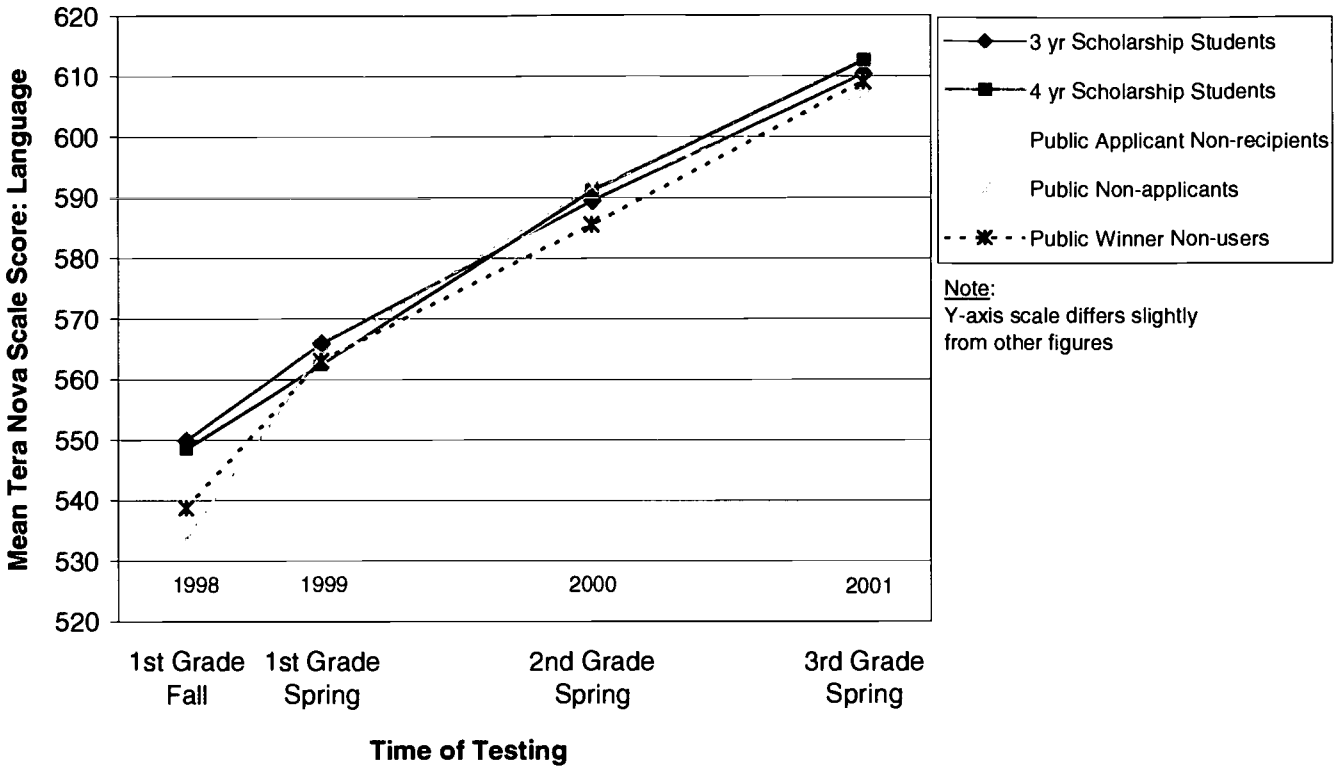


Figure 4. Mathematics Achievement: 1998 to 2001

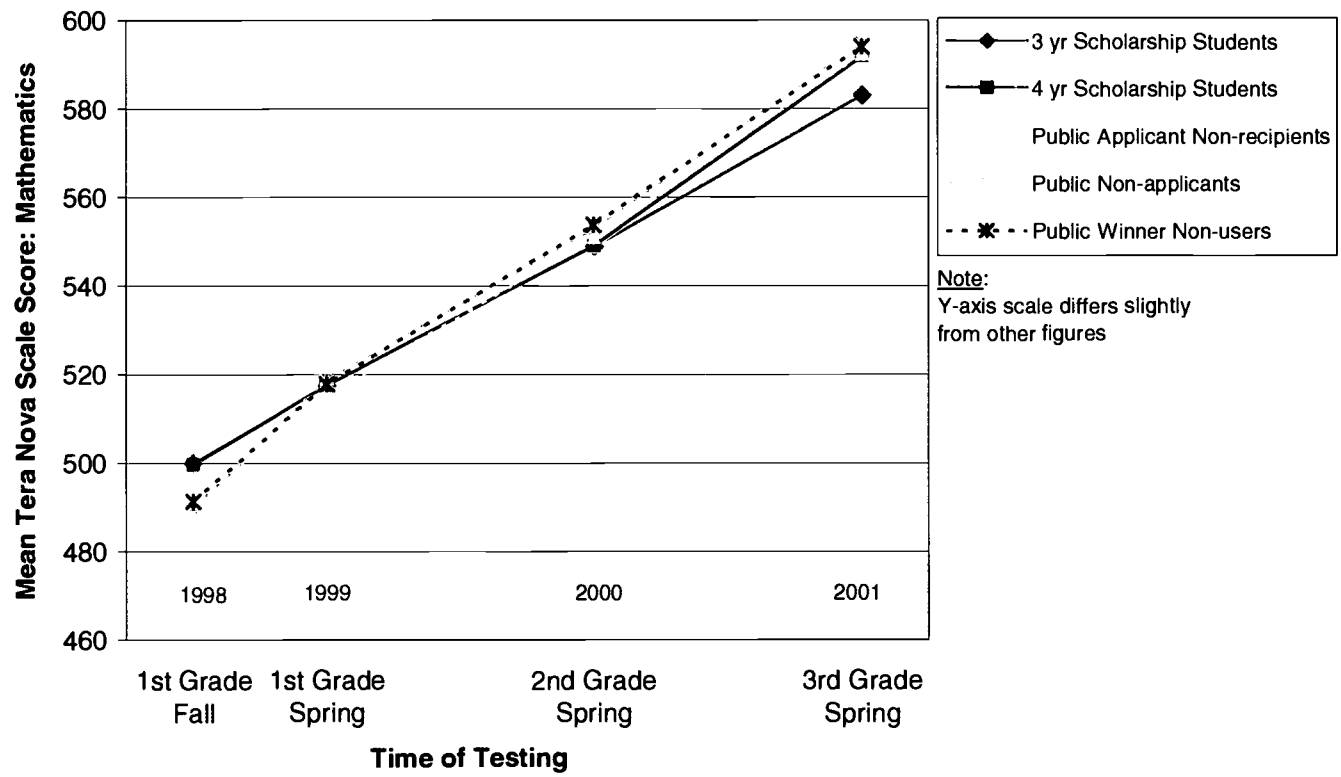
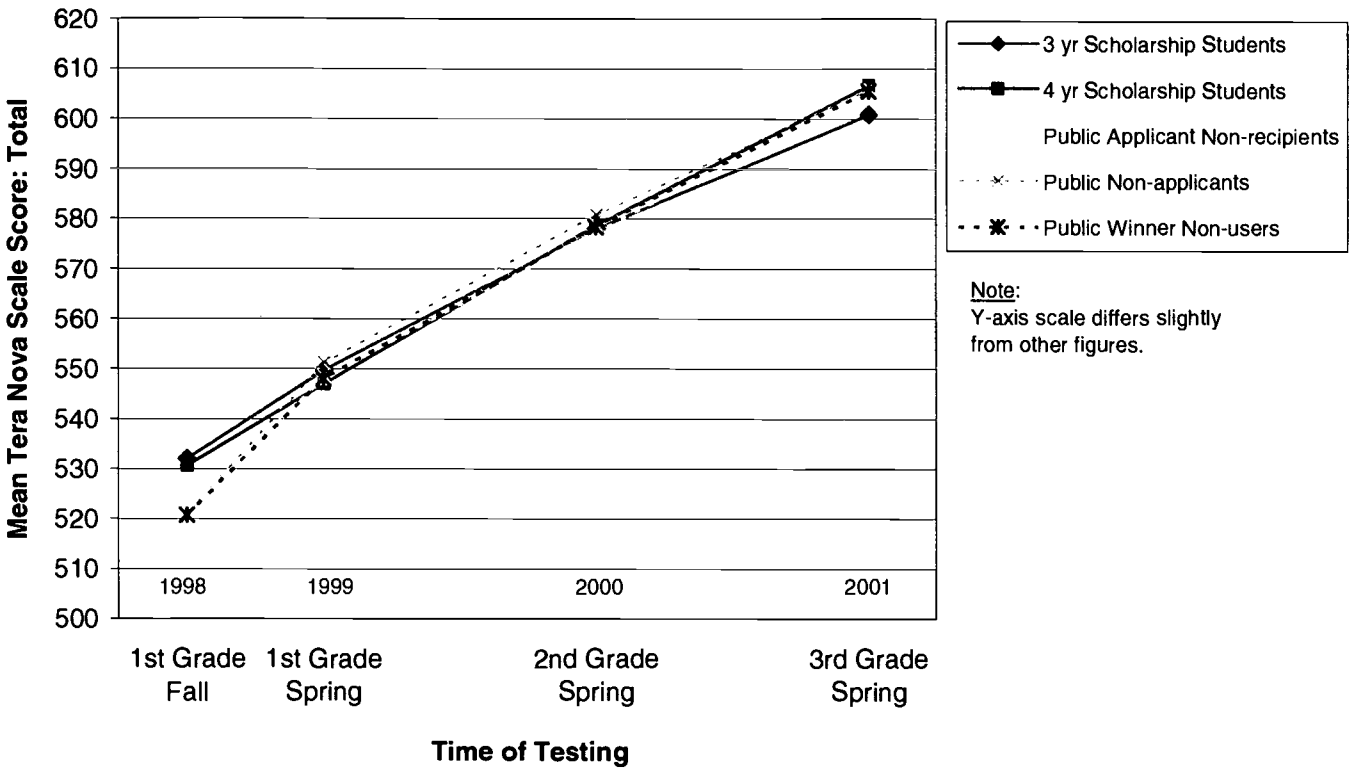


Figure 5. Total Achievement (Average of Reading, Language, and Mathematics): 1998 to 2001



Summary

Across the evaluation findings from these early years of students' experience in the CSTP, there is little that is provocative. Scholarship students enter the program with somewhat different academic and demographic characteristics than students who attend public schools. Public school teachers are more likely to have completed graduate coursework, particularly up to the masters' level, than are private school teachers. Class size and teacher experience, at least for the samples of classrooms we examined, were quite similar between public and private schools. Furthermore, student academic achievement, a factor watched closely by those on both sides of the voucher issue, presents no clear or consistent pattern than can be attributable to program participation.

One finding that emerged, however, should not be overlooked. By the end of third grade in 2001, the current cohort of scholarship students, as a whole, is comprised of fewer minority students and their families tend to have higher incomes in comparison to their peers in pubic schools. More important, however, is the finding that this difference in the

ethnic/racial and socioeconomic composition of the groups likely is a result of the students who enter the CSTP from private schools after the start of each academic year. Students who entered the CSTP from public schools tend to accept the award prior to the beginning of the school year, and they are statistically equivalent to their classmates who remained in public schools (approximately 80% minority and 87% from families eligible for free lunch). In contrast, students entering from private schools tend to be offered and accept the award after the beginning of the school year, and they differ significantly from their peers in public schools in terms of minority status and family income (approximately 46% minority and 54% from families that would meet the Federal requirements for free lunch eligibility).

Therefore, it seems that the random lottery, held by the CSTP office each spring, generally serves to award scholarships to students who have been in public schools and who are demographically similar to their peers in Cleveland public schools. However, as the beginning of the academic year approaches and then passes, some students elect not to use scholarships that they were awarded. These scholarships are then offered to other applicants to the program (who may not have been selected in the initial lottery). Students who, at this later point, are awarded and choose to use a scholarship tend to be Caucasian, from families with higher incomes, and enrolled in private schools at the time of the award.

The findings and the conclusions that we have attempted to draw are preliminary. They represent only the first of what are to be multiple sets of findings drawn on students over a period of several years. As a result, the findings are incomplete and understandably inconclusive. Yet, information can be gained about the ways in which a choice program, or schools more generally, influence the early school experience of children.

Closing Thoughts and Future Directions

The recent ruling by the U.S. Supreme Court upholding the constitutionality of the Cleveland Scholarship and Tutoring Program has removed a cloud of doubt about the program's future that has loomed over it since its inception. Many analysts and researchers now speculate that other states will propose new voucher programs, and that existing programs, particularly Cleveland's, will now thrive. Even opponents of such programs have increased their calls for sound, empirical research on these programs that can inform the re-ignited policy debate.

The longitudinal study that is being conducted in Cleveland represents the longest running evaluation of a publicly-funded voucher program that has ever been conducted. As such, it affords a unique opportunity to examine the characteristics, operation, and impacts of such programs over time. The data that have been collected thus far, and that are reported above, represent work done through spring of 2001. However, the evaluation continues.

At the present time, two key additional data collection activities have already taken place. Students in the cohort, who were enrolled in fourth grade during the 2001-02 year, were administered the standardized achievement test by the evaluation team in April of 2002. These tests are currently being scored and these data will be integrated into the existing data set. As a result, we will shortly have available an additional year of achievement data for these students. Furthermore, plans are underway to conduct testing of the cohort as they complete fifth grade in the spring of 2003.

In addition, telephone interviews were conducted with random samples of parents and guardians of both public and private school students in Cleveland during late spring and early summer of 2002. These extensive interviews included a range of critically different groups of families, each of which possesses unique attributes which may influence their perceptions of schools and school choice. Sampling across grades 2, 4, and 8, interviews were conducted with families who:

1. Had been awarded and were using scholarships to send their children to private schools (representing what we have called *winner users),*
2. Had applied for, but were not awarded scholarships and had enrolled their children in public schools (representing what we have called *applicant non-recipients),*
3. Had applied for and were awarded scholarships, but who chose instead to enroll their children in public schools (representing what we have called *winner non-users*),
4. Had never applied for scholarship, and who had enrolled their children in public schools (what we have called *non-applicants)*, and importantly,
5. Had chosen to enroll their children in charter schools (referred to as community schools in Ohio).

Data drawn from these interviews will provide insights into the criteria by which families make decisions about their children's education. The inclusion of families of charter school students and the explicit inclusion of scholarship winner non-users represents a new approach to this type of study of school choice.

The recent Supreme Court ruling and the renewed importance of rigorous research on this issue raise a number of new questions. For example, will applications to the Scholarship and Tutoring Program in Cleveland increase dramatically because the stability and future of the program are now secure? Anecdotal evidence from past years suggests that some families elected not to pursue or use a scholarship because of the program's uncertain legal status. Relatedly, will private schools respond to the ruling by creating more classrooms for scholarship students, or will new private schools open or elect to participate in the program? Will the characteristics of families who apply for scholarships change? It is not unreasonable to believe that previous uncertainty about the program may have led to a disproportionate number of applicants to be families whose children already attended private schools. These families would not have to worry about losing their scholarship if the program were discontinued. In contrast, many families whose children were in public schools may have been interested in a scholarship for private school enrollment, but chose not to risk the disruption of moving their children back into public schools if the program was discontinued.

What are the long-term impacts of vouchers on schools, families, or students? Answers to this question and a range of others can be found. The ongoing work in Cleveland is one step toward this goal, as is the work of other researchers in other parts of the country. But it will take time, careful study, and patience to obtain the evidence that will be required by policy makers and families as the future of school choice unfolds.

THIS DOCUMENT CONCERNS THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM AS ESTABLISHED PURSUANT TO OHIO REVISED CODE SECTION 3313.975, AND SHOULD NOT BE CONFUSED WITH THE CLEVELAND SCHOLARSHIP PROGRAM, INC.

EVALUATION OF THE CLEVELAND SCHOLARSHIP AND TUTORING PROGRAM

TECHNICAL REPORT
1998 - 2001

Kim K. Metcalf, Ph.D.
Director

Stephen D. West, M.A.

Natalie Legan, M.S.

Kelli Paul, M.S.

William J. Boone, Ph.D.

October, 2002

Indiana Center for Evaluation

509 East Third Street
Bloomington, IN 47401-3654
http://www.indiana.edu/~iuice/

Kim K. Metcalf, Ph.D.
Director

812-855-4438
800-511-6575
Fax:812-856-5890
iuice@indiana.edu

Table of Contents

List of Tables

List of Figures

1.0 Introduction

In the autumn of 2002, the Cleveland Scholarship and Tutoring Program (CSTP)[1] begins its first year free of the legal uncertainty that has surrounded the program since its inception. In the weeks following the Supreme Court's ruling that the program is constitutional, attention has been refocused on the ongoing evaluation of the program. As the longest running evaluation of any publicly-funded voucher program to date, and as the only fully longitudinal study of the impacts of vouchers on students over time, the state-contracted evaluation provides a unique opportunity to enhance understanding of voucher and other choice programs.

The present report is based upon data collected during the period 1998-2001, the first three years of the second phase of the evaluation. It is the second report to be released during phase two.[2] Although the longitudinal study includes examination of a range of factors associated with the CSTP (e.g., impacts on schools and families; characteristics of students, teachers, and schools; etc.), a focus of the study has been on the impacts of participation in the CSTP on students' academic achievement. Using multiple comparison groups and annual measurement of students' academic performance, the study has examined both the level of achievement and achievement growth from the beginning of first grade through the end of third grade.

Although different evaluation questions have been emphasized over the years of the evaluation, the portion of the study reported here was guided by the following questions about the program, its operation, its impact, and the implications it has for voucher programs in other areas:

1. What are the characteristics of students who participate in the CSTP, and how do they compare with: (a) students who did not apply to participate; (b) students who applied to participate but did not receive an award; (c) students who applied to participate, received an award, but who chose to not use the award; and (d) students who received and used a scholarship in the past, but who chose to withdraw from the program and attend pubic schools?

[1] The present evaluation and report concerns the Cleveland Scholarship and Tutoring Program, as established pursuant to Ohio Revised Code Section 3313.975, and should not be confused with The Cleveland Scholarship Programs, Inc. – a privately funded organization.

[2] All prior reports, from both phases one and two, are available through the Ohio Department of Education or through the Indiana Center for Evaluation (www.indiana.edu/~iuice).

2. What are the characteristics of the classrooms and teachers with whom scholarship students work in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?

3. What is the impact of participation in the CSTP on student academic achievement in comparison to students who do not participate in the program for various reasons (e.g., students who applied for but did not receive a scholarship; students who applied for, received, but subsequently chose not to use a scholarship; student who withdrew from the program to attend public schools; and public school students who never applied for a scholarship)?

The report is organized into three sections. A first section describes the evaluation approaches and methodologies that were used during the most recent year of the project (2000-2001). Section two presents detailed results of the data analyses and is organized around the three evaluation questions described above. The final section discusses current results in the context of earlier findings, presents conclusions from the study to date, and outlines the intended future activities of the evaluation.[3]

1.1 Phase Two of the Cleveland Scholarship and Tutoring Program Evaluation

The current phase of the project was initiated during the 1998-99 academic year. This phase of the study is designed to include several methodological characteristics that have been lacking in earlier research on vouchers. Specifically, the current phase: (1) focuses on a cohort of children from the time they began their education in kindergarten or first grade either in private or public schools; (2) uses multiple comparison groups of children (e.g., scholarship users, public school non-applicants, applicant non-recipients, etc.); (3) longitudinally examines the impact of the program on families, children, and schools; and (4) tracks the performance and academic growth of individual students over time.

Autumn of 1998 marked the start of the current phase with collection of achievement data on 780 first grade scholarship students, 541 first grade public school applicant non-recipients, and 1,233 first grade non-applicants. Since that time, achievement data have been collected from this

[3] The approaches and methodologies from previous years have been reported in previous reports (see Metcalf, K.K., *Cleveland Scholarship and Tutoring Program Evaluation: 1998-2000 Technical Report*, Bloomington, Indiana; Indiana Center for Evaluation, 2001).

cohort each spring. Therefore, the current report presents findings from autumn, 1998 (early first grade) through spring of 2001 (late third grade).

2.0 Evaluation Methods and Approaches

The first two years of the current phase of the evaluation (1998-1999 and 1999-2000) primarily focused on identifying and tracking the target sample population and on collecting baseline demographic and achievement data on the target students during their first and second grade years. The present report includes the addition of another year of achievement data (i.e., from third grade, 2000-2001) and outlines the evaluation activities planned for the 2002-2003 school year, which – in addition to the continued collection of achievement data – will emphasize the impact of the program on teachers, schools, and families.

A mixed-model research design has been used to address the three previously defined primary research questions. Multiple comparison groups have been distinguished and used in an attempt to delineate program factors and effects. The primary data source used to addresses student achievement is the Terra Nova, a standardized test produced by CTB/McGraw-Hill. Each year, the Terra Nova has been administered to targeted students by representatives of the Indiana Center for Evaluation. At the time of test administration, trained proctors also collect classroom-level data through teacher interviews and limited classroom observations from each school in which a test is administered. Additionally, data have been collected from Cleveland Scholarship and Tutoring Program office records as well as Cleveland Municipal School District (CMSD) records.

2.1 Data Sources and Collection

For this report, one primary source, three supplemental sources, and one ancillary source of data were used. The Terra Nova, providing student academic achievement measures, is the primary data source, whereas Scholarship and Tutoring Program records, CMSD records, and interview/observational records of characteristics of classrooms and teachers serve as supplemental data sources. The ancillary data source consists of information on student mobility over the three years covered by this report, which was obtained by examining records maintained by the evaluation team at the Indiana Center for Evaluation.

Additionally, new records of extant data were collected from the CSTP office and the CMSD office, thereby enabling an entire record audit to be conducted by the evaluation team. The audit

enabled the further identification and classification of students within our target population, which has improved the reliability (accuracy, completeness, etc.) of our sample.

2.1.1 Achievement Test Scores

Throughout phase two of the evaluation, student academic achievement is measured using the Terra Nova (CTB, McGraw-Hill). This achievement test was selected early in the project because: (a) none of the schools in which data are collected use the Terra Nova as their primary off-grade testing tool, (b) the test is among the most progressive in its use of contemporary testing principles and formats, and (c) scale scores are provided that allow comparisons across time. In the first and second grade, the selected versions of the test (Levels 10, 11, and 12) provided scores for each student in reading, language, mathematics, and a total score – which is the average of the other three achievement measures. In the third grade, the version of the test selected (Level 13) included science, and social studies in addition to reading, language, mathematics, and a total score.[4]

Baseline data were collected from first grade students using Level 10 of the Terra Nova, which was administered during the fall of 1998. Achievement data again were collected from first grade students in the spring of 1999 using Level 11 of the Terra Nova. Since that time, the Terra Nova has been administered each spring. Level 12 was administered to second grade students during the spring of 2000, and Level 13 was administered to third grade students during the spring of 2001.

For test administration, proctors are trained and supervised using standardized administration procedures. Proctors administer the test to target students in public and private schools over two testing sessions. The first testing session, which covers the reading and language portions of the test, is administered in the afternoon. The second testing session, which covers the mathematics, social studies, and science portions of the test, is administered on the following morning. A proctor to student ratio of 2:20 is used as a baseline, with additional proctors provided when large classrooms are used to test multiple groups of students.

[4] The total score on the Level 13 test is calculated by CTB-McGraw Hill using the Language, Reading, and Mathematics scores (i.e., the average of the three scores). The total score on all Terra Nova tests is calculated in a similar manner, regardless of level.

The longitudinal nature of the current project requires the use of linear scaled scores, rather than the more commonly used, but non-linear, normal curve equivalent (NCE) scores. Scale scores on the Terra Nova range from approximately 100-900 and, unlike non-linear scores (e.g. grade equivalent or percentile rank scores), can be compared and combined across multiple years of testing. In particular, the scale scores on the Terra Nova can describe student achievement on a continuum from kindergarten through 12th grade. Although the entire scale ranges from 100-900, each level (e.g., the level appropriate for third graders) has a unique minimum and maximum attainable score, such that the overlapping levels, which increase in difficulty, are linked to span the entire continuum from 100-900 (representing achievement across grades K-12). Thus, as the grade level of test difficulty increases, the range of attainable test scores, by design, increases as well.[5]

2.1.2 Scholarship and Tutoring Program Office Records and Public School Records

Cleveland Scholarship and Tutoring Program (CSTP) office records are the primary source of extant data for the evaluation. The records serve as a means of identifying students as scholarship winner-users, applicant non-recipients, and scholarship winner-non-users. Through the process of elimination, these records also enable the evaluation team to identify non-applicants. Furthermore, by tracking this information across time, former scholarship winner-users can be identified (i.e., students who applied for and used a scholarship in the past but, at some point in time, returned to public schools). More specific information about the various sub-populations of the sample will be further described in Section 2.2 of this report, *Sample and Sample Selection.* Additionally, CSTP records provided information on the names of the private schools in which scholarship students are enrolled, thereby enabling the evaluation team to track students' schools of attendance and school changes across time. These records are updated as scholarships are awarded and subsequently updated throughout the year if families change schools.

[5] For example, scale scores on the Level 10 Reading section of the Terra Nova range from a minimum of 355 (zero items correct) to 626 (20 items correct), scale scores on the Level 11 Reading section range from a minimum of 407 (zero items correct) to 701 (25 items correct), and scale scores on the Level 12 Reading section range from 423 (zero items correct) to 722 (32 items correct). The test is designed so that a student who receives a score of 600 on the Level 10 Reading section will also score close to 600 on the progressively more difficult Levels 11 and 12 – plus or minus a degree of measurement error. That is, a score of 600 represents the same level of achievement/mastery of the material and concepts regardless of the level administered (within the limits of the floor and ceiling of a given test level).

CSTP records are also used to obtain demographic information on students whose families applied for a scholarship at some point between the time their child entered kindergarten and third grade. The demographic data provide descriptive information and allow for comparisons to be made among participating, non-participating, and formerly participating students. Specifically, CSTP records provide demographic information on student race, sex, family size, and family income.

Cleveland Municipal School District records also are used for purposes similar to CSTP records. That is, this data source provides demographic information on students as well as school of enrollment for students whose most current school is not available from CSTP records (e.g. former scholarship users, applicant non-recipients, and/or winner-non-users). Demographic information provided through the public school dataset includes student race, sex, and meal code status (i.e., eligibility for free or reduced price lunch). When possible, these data are used in conjunction with the CSTP records. For public school non-applicants, who have no information in CSTP records, this is the only demographic information available.

Family income often has been found to be strongly related to students' academic achievement and, thus, was included as one variable in our earlier evaluation work. It was similarly believed important to include some estimate of this variable in the present phase of the evaluation. However, no single measure of family income is available for all students. All scholarship applicants are required to report their family income to the CSTP office, and this information is verified when vouchers are awarded. Therefore, CSTP office records include family income data on most of the present participating sample. Public school records, however, provide only free or reduced price lunch eligibility data for students attending public schools. No linearly scaled measure of family income is available for public school students.

To create a single estimate of family income for the entire sample, the Federal Child Nutrition Programs Income Eligibility Guidelines (used to determine eligibility for free and reduced price meals for students across the country) were utilized in the present investigation to calculate an estimated meal code for students in the sample with reported family income. In this manner, a single, categorical proxy measure of family income was created for as many students in the sample as possible. Students whose families reported income information (e.g., all scholarship

recipients and most scholarship applicants) were assigned a code that designated them as eligible for: (1) free lunches, (2) reduced price lunches, or (3) not eligible for assistance (lunches are paid by students' families). These are the same meal code categories reported for the public students in the sample. Based on Federal Guidelines, the estimated meal codes were calculated according to the schedule displayed in Table 1.[6]

Table 1. Estimated Meal Code Calculation Guidelines

Household Size	Estimated Meal Code: Eligibility Guidelines (Based on *Annual Family Income* and *Household Size*)		
	1 Free Meals	2 Reduced Meals	3 No assistance / Paid
1	$10,855	$10,856 – $15,448	> $15,448
2	> $14,625	$14,626 – $20,813	> $20,813
3	> $18,395	$18,396 – $26,178	> $26,178
4	> $22,165	$22,166 – $31,543	> $31,543
5	> $25,935	$25,936 – $36,908	> $36,908
6	> $29,705	$29,706 – $42,273	> $42,273
7	> $33,475	$33,476 – $47,638	> $47,638
For each additional family member add:	$3,770	$3,770 – $5,365	$5,365

The estimated meal code categories calculated for private school students match the actual meal code categories reported for students in public schools. If neither 2000-2001 family income nor 2000-2001 meal code information was available for a particular student, the student's most recently reported family income or meal code status was included in the estimated meal code variable for this report. This process allowed the evaluation team to assign an estimated or actual meal code to nearly all students in the sample, whether or not they had attended public schools or had applied for a scholarship. It should be noted, however, that these estimated meal codes are not exact measures of family income.[7]

[6] The Income Eligibility Guidelines to determine eligibility for free and reduced price school meals are released by the U.S. Department of Agriculture, Food and Nutrition Service (FNS). The Guidelines used to calculate an estimated meal code for the present investigation were effective from July 1, 2000 to June 30, 2001. The Income Eligibility Guidelines can be obtained from the Federal Register: April 4, 2000, Vol. 65, No. 65, which is available online at http://www.gpo.gov/su_docs/aces/aces140.html

[7] Simple correlations between this estimated meal code and both the *assigned meal code* and *reported family income* are r=.98 and r=.79, respectively.

Data maintained by the evaluation team over the course of this evaluation comprise an additional source of demographic data. Data on each targeted student's school of enrollment from kindergarten through third grade provides a basic indicator of student mobility across time. Specifically, student mobility was calculated by tracking the number of times that each student changed schools from kindergarten (1997-1998) to third grade (2000-2001). Therefore, the student mobility indicator ranges from zero (no school changes/ one school attended from grade K-3) to three (the student changed schools every year from grade K-3). Because there are instances of students in the sample attending one school, moving to a different school, and then changing back to attend their original school (i.e., they made two school changes but only attended two different schools), the *number of school changes* made by students from kindergarten to third grade was used in the analyses of student mobility as opposed to the total number of schools attended during this time period.

There are inconsistencies, however, in the availability of student data due to the varying record keeping systems of the CMSD office and the CSTP office. These discrepancies between public and private school records make it challenging to create consistent demographic variables. Nonetheless, some demographic data are available for each student in the sample. These variables include the following:

Variable	Codes
Student Sex	1 = Male, 2 = Female
Student Race/Ethnicity [8]	1 = non-minority, 2=minority
Estimated Meal Code	1 = free, 2 = reduced, 3 = full pay
Student Mobility	0 = no school changes, 1 = one school change, 2 = two school changes, 3 = three school changes
School of enrollment [9]	1= one school from K-3, 2 = two schools from K-3, 3 = three schools from K-3

[8] As was done in the past, race was coded dichotomously as minority and non-minority (i.e. Caucasian) due to small numbers of students in categories other than African-American and Caucasian classifications.

[9] School of student enrollment was coded separately in an attempt to differentiate between the number of moves a student made over time and the number of different schools a child attended over time. For example, a child who attended school A, then moved to school B, then moved back to school A, made two school changes and attended two different schools, whereas a child who attended school A, then moved to school B, and then moved to school C, made two school changes and attended three different schools.

2.1.3 Classroom Surveys

Brief classroom surveys conducted by proctors on the testing days in each school comprise a third primary data source. Guided questionnaires are used to obtain information regarding the teachers of targeted students and the classrooms in which targeted students are primarily situated. Specifically, information is obtained on the teacher and classroom for each targeted student relating to the following: (a) class size, (b) total years of teaching experience, (c) years of teaching experience at current school, (d) teacher's level of education, and (e) whether or not the teacher is certified. The level of education (i.e., teachers' highest degree earned) is coded as follows: 1 = No degree, 2 = BA/BS, 3 = BA/BS +, 4 = MA/MS, 5 = MA/MS +, 6 = Ed.S, 7 = ABD, 8 = Ph.D. These data are used descriptively to provide information specific to the classroom context in which the students learn.

Table 2 presents each type of data collected and the source from which it was obtained. The data described above were used independently and in combination in all subsequent analyses.

Table 2. Data Types and Sources

Type of Data	Public School Records	CSTP Office Records	Terra Nova	Evaluation Office Records	Classroom Interviews
Student Sex	X	X	X		
Student Race	X	X			
Family Size		X			
Family Income		X			
Meal Code	X				
Estimated Meal Code				X	
Academic Achievement			X		
School of Enrollment	X	X			
Class Size					X
Teacher Experience-Total					X
Teacher Experience at present school					X
Teacher Education					X
Teacher Certification					X
Student Mobility				X	
School(s) of Enrollment				X	

2.2 Sample and Sample Selection

Research and evaluation of voucher programs has generated substantial debate over the most appropriate comparison group(s) against which to assess program impacts.[10] In the present study, multiple comparison groups are used to ensure that the most complete and representative judgments can be made. To initiate the longitudinal evaluation, the evaluation team obtained the broadest possible sample of participating and non-participating students during their first grade year in 1998-1999. These students subsequently have been, and will continue to be, followed over a multi-year period that currently includes first grade (1998-1999), second grade (1999-2000), and third grade (2000-2001).

Over the three-year period covered by the present report, evaluation activities have been designed to identify and follow students in each of five groups. These groups include: (a) *scholarship students* attending private schools; (b) *scholarship applicant non-recipients* – students who applied for but did not receive a scholarship and who attend public schools; (c) *scholarship winner non-users* – students who applied for and received a scholarship but did not use the scholarship and attend public schools; (d) *former scholarship recipients* – students who received and used a scholarship for one or more years, subsequently withdrew from the CSTP, and now attend public schools; and (e) *non-applicants* – public school students whose families never applied for a scholarship. Although private school students could also be classified according to this taxonomy, it should be emphasized that, for the purpose of the present evaluation, scholarship applicant non-recipients, scholarship winner non-users, former scholarship recipients, and non-applicants are comparison groups comprised solely of students who attend public schools.

[10] For example, William G. Howell and Paul E. Peterson, *The Education Gap: Vouchers and Urban Schools* (Washington, DC: Brookings Institution, 2002); Martin Carnoy, *Do School Vouchers Improve Student Performance?* (Washington, DC: Economic Policy Institute, 2001); Kim K. Metcalf, Ronald R. Beghetto, and Natalie A. Legan, *Voucher Research: Understanding the Methodological Dilemmas* (Paper presented at the annual meeting of the American Educational Research Association, 2002). Howell and Peterson argue that the most appropriate approach to examining the impact of vouchers is through randomized field trials (RFTs), in which eligible applicants are randomly assigned (e.g., through a lottery process) either to receive a voucher or not. Carnoy first speculated that such an approach might overestimate the impact of vouchers due to a *negative* treatment effect for those who wanted, but did not receive a voucher. Metcalf, Beghetto and Legan draw on the work of other voucher researchers to build a case that such an effect may exist and, as a result, that multiple comparison groups are required to fully understand the impacts of vouchers.

Scholarship students, applicant non-recipients, scholarship winner non-users, and former scholarship students were first identified through CSTP office records. These records also indicated scholarship students' private schools of enrollment. Public schools of enrollment, however, were not always available in the dataset for applicant non-recipients, scholarship winner non-users, and former scholarship students. Public school records and private school contacts were used in an attempt to determine the school of enrollment for students in these three groups.

The final comparison group consisted of public school students whose families had never applied for a scholarship. This group of students was identified in the first year of this phase of the evaluation as testing was conducted in every public school and all participating private schools in Cleveland. The public school non-applicant group represents students who were enrolled in public school classrooms during first grade (1998-99) with other students identified as applicant non-recipients, scholarship winner-non-users, and former scholarship recipients (who withdrew from the program after kindergarten). To provide the most comfortable and consistent environment for testing, and to provide schools with an additional set of complete classroom achievement data, most schools elected to test all students in classrooms, even if they had not been identified as a target student.[11] Consequently, in the process of testing scholarship applicant non-recipients, scholarship winner non-users, and former scholarship recipients, who were attending public schools, a large number of non-applicant public school students also were tested. These students came to constitute the fourth group with which scholarship students were compared.

To ensure that each student's involvement with the CSTP is properly identified and tracked across the multi-year study, the present evaluation utilizes a classification system designed to capture the most important information about the students' scholarship status, and changes in their status, across the course of the investigation. Specifically, the following hierarchical coding scheme was utilized (codes are listed in descending order, such that each code overrides all codes below it):

[11] The Cleveland Municipal School District has since changed this policy and now allows testing only of specifically targeted students.

1. Active scholarship students are identified as *scholarship recipients*, regardless of their status prior to receiving and using the scholarship. For example, the group of scholarship recipients in third grade (autumn, 2000) is comprised of students who used a scholarship as a third grader, and who entered the program in kindergarten, first grade, second grade, or third grade. Students who entered later in their schooling (e.g., in second or third grade) may have previously applied for but not received a scholarship, or never done so (as well as other combinations of prior scholarship status codes).

2. Students who have been confirmed as scholarship recipients and users in one or more previous years, but are not currently listed as a scholarship recipient are classified as *former scholarship recipients*, regardless of their status prior to or following their withdrawal from the CSTP (unless they re-enter the program, in which case their scholarship recipients status overrides their former scholarship recipient status). For example, students who participated in the CSTP for one school year and withdrew after kindergarten are classified as former scholarship recipients even if they subsequently applied for and did not receive another scholarship (i.e., these students are not identified as applicant non-recipients).

3. Students who are listed as scholarship winners in a given school year but did not use the scholarship during that time period are classified as *scholarship winner non-users*, regardless of their scholarship status prior to or following the year in which they received but did not use a scholarship – unless, of course, they subsequently receive and use a scholarship (#1 above) or use a scholarship and withdraw from the program (#2).

4. Students who do not qualify for any of the aforementioned scholarship status code designations *and* whose families have applied for but did not receive a scholarship at any point from kindergarten to third grade are classified as *scholarship applicant non-recipients* – regardless of whether or not their families subsequently apply for and do not receive a scholarship again. For example, students identified as non-applicants in kindergarten and applicant non-recipients in first grade are identified as applicant non-recipients, even if their families do not apply for a scholarship in second and third grade.

5. Finally, students who have never applied for a scholarship at any point from kindergarten to third grade are classified as *non-applicants*.

The coding process for this study began in autumn, 1998, when students were beginning their first grade year, and it was repeated and updated in spring, 1999, spring 2000, and spring 2001 (i.e., as students were tested at the end of first grade, second grade, and third grade, respectively). Through this process of updating student status, and because students move into and out of the various groups, sample sizes vary across the four testing episodes. Furthermore, the *target sample* (i.e., the intended sample of students classified through records of earlier testing into one of the five target groups) and the *actual sample* (i.e., the resulting sample of students present during the two-day testing episodes and for whom achievement and demographic data actually

are obtained) are not identical at each testing period. Due to student transience, double promotion of target students, retention of some target students, absences during the testing episodes, inconsistent school records, and other uncontrollable and unintended factors, the actual sample obtained from each testing episode differs from the target sample. Over the period covered by the present report, the target and actual samples are presented in Table 3.

Table 3. Target and Actual Samples

Student Group	Time of Testing							
	Autumn, 1998 (early 1st Grade)		Spring, 1999 (late 1st Grade)		Spring, 2000 (late 2nd Grade)		Spring, 2001 (late 3rd Grade)	
	Target	Actual	Target	Actual	Target	Actual	Target	Actual
Scholarship Recipients/Users	883	787	789	779	695	647	672	671
Public Applicant Non-recipients[a]	480	452	467	423	436	362	377	355
Public Non-Applicants	–	1233	1436	1245	1041	934	1183	1195
Public Winner Non-Users[b]	80	68	79	69	89	71	96	95
Former Scholarship Recipient/Users – Public[c]	32	22	51	49	97	72	124	112

General Notes: (1) The "Target" sample includes students for whom the evaluation team possessed current school information and, as a result, intentionally scheduled proctors to administer the standardized test at those schools.

a. In previous reports, the number of *Applicant Non-recipients* included both public and private school students. The numbers reported in the present report, however, include only public school *Applicant Non-recipients,* as they comprise the comparison group of interest.

b. Data presented in previous reports did not distinguish between *Applicant non-recipients* and *Winner non-users* during the first grade year. The numbers presented in this report for *Winner Non-users* during the first grade testing episodes were calculated retrospectively based on the evaluation team's audit of CSTP office records during the 2000-2001 and, therefore, may differ from numbers reported in previous reports.

c. Previously, *Former Scholarship Recipient/Users* were included in the *Winner Non-user* category. The numbers presented in this report prior to the third grade testing episode were calculated retrospectively based on the evaluation team's audit of CSTP records during the 2000-2001 school year and, therefore, may differ from numbers reported previously.

2.2.1 Subgroups

The longitudinal design of this evaluation also allows some of the groups discussed above to be further divided based on the students' unique patterns of scholarship status across the four testing episodes.

Scholarship Recipients: Four groups of scholarship recipients have been identified based on their entry into the CSTP (ignoring scholarship status prior to entering the program). These scholarship recipient groups are: (a) four-year scholarship recipients – students who entered the program in kindergarten (autumn, 1997) and have participated in the program for four school

years, from kindergarten though third grade (2000-2001); (b) three-year scholarship recipients – students who entered the program in first grade (autumn, 1998) and have participated in the program for three school years, from first grade through third grade; (c) two-year scholarship recipients – students who entered the program in second grade (autumn, 1999) and have participated in the program four two years, in second and third grade; and (d) one-year scholarship recipients – students who entered the program in third grade (autumn, 2000) and have participated for one school year.

The groups of scholarship recipients who entered the program at various points from kindergarten (1997-1998) to third grade (2000-2001) have been identified to determine whether and how students who entered the CSTP later (e.g., in second or third grade) differ from students who entered the program earlier (e.g., in kindergarten or first grade). In general, the four-year scholarship recipients constitute the primary group of interest in the present evaluation because they have continuously participated in the program since kindergarten and have never attended public schools.

Former Scholarship Recipients: Former recipients are comprised of subgroups of students who participated in the program for varying amounts of time and withdrew at different points between their enrollments in kindergarten through third grade. Three groups of former scholarship recipients have been identified and classified based on when they withdrew from the CSTP: (a) three-year former scholarship recipients – students who participated in the program for one school year, during kindergarten (1997-1998), withdrew from the program after kindergarten, and have attended public schools for *three years* (first grade through third grade); (b) two-year former scholarship recipients – students who participated in the program for two years, during kindergarten (1997-1998) and first grade (1998-1999), withdrew from the program after first grade, and have attended public schools for *two years* (second grade and third grade); and (c) one-year former scholarship recipients – students who participated in the program for three years, from kindergarten through second grade (1999-2000), withdrew from the program after second grade, and have attended public schools for *one-year* (during third grade). The former scholarship recipient subgroups, classified based on differential exit from the CSTP, were identified to investigate whether and how students who withdrew from the program differ as a function of their date of exit.

Applicant Non-recipients and Scholarship Winner Non-users: The scholarship applicant non-recipients and scholarship winner non-users included in the present sample were identified based on the coding process described above. As a result, these groups include students who may or may not have been identified as members of their current groups across all of the testing episodes. For instance, the 2000-2001 third grade applicant non-recipient group not only includes students whose families applied for and did not receive a scholarship in kindergarten, but it also includes students whose families applied for and did not receive a scholarship in first grade and in second grade, as well as students who have applied for a scholarship every year (among the many other possible combinations). In other words, once a student has been identified as an applicant non-recipient at any point in time, the student's scholarship status carries forward, regardless of whether or not that student applies for and does not receive a scholarship again. Scholarship winner non-users can be classified and subdivided in a similar fashion. Preliminary statistical analyses of the achievement scores, demographic characteristics, and classroom characteristics, however, failed to reveal statistically significant differences among the applicant non-recipient subgroups and among the winner non-user subgroups. Therefore, the applicant non-recipient and winner non-user subgroups have been aggregated and treated as two separate homogeneous groups of students, namely applicant non-recipients and winner non-users.

2.2.2 Sample Selected for 2000-2001

Following the spring 2001 (late third grade) testing episode, achievement and demographic data have been obtained from four testing episodes (early first grade, late first grade, late second grade, and late third grade). As mentioned above, due to the inherent difficulties involved with tracking students across multiple years in a longitudinal study, the actual sample (as of spring 2001) does not contain complete achievement and demographic data from all of the students targeted during testing. That is, students in the sample exhibit varying amounts of data across the four testing episodes.

With longitudinal data, however, even moderate amounts of missing data on each variable can result in very few student records with complete data on every variable. Because standard statistical packages handle missing data by excluding incomplete cases from the analyses (i.e., listwise deletion), a substantial portion of the data often is omitted. For instance, discarding students who are missing only one of four possible achievement scores on a given measure can

be very inefficient and result in a considerable loss of statistical power. Unfortunately, standard repeated measures or mixed-design analyses of variance do not include such student records, and the three time periods for which information was collected are ignored. Furthermore, unless the omitted student records (due to missing data) comprise a completely random sub-sample of the data set, the discarded records may systematically differ from the complete records and parameter estimates may be seriously biased.

To address the issue of missing data, a multiple imputation technique[12] was utilized to replace the missing achievement scores with estimated (i.e., mathematically simulated) values randomly drawn from a predictive probability distribution of missing scores and error terms. The predictive probability distributions used for imputing missing values were mathematically derived from the linear relationship between the observed scores on a given variable and the other variables in the data set (e.g., other achievement scores, demographic variables, and classroom characteristics). A detailed discussion of the multiple imputation procedure is presented in Appendix A.

2.2.3 The Imputed Sample

With the exception of analyses addressing the question of differential entry into the CSTP, all of the analyses of achievement data presented in this report have been conducted on the single data set that resulted from combining the ten complete imputed (simulated) data sets – with all of the missing data replaced by estimated values.[13] The resulting imputed data set is comprised of 1595 students out of the 2302 students identified as members of one of the five targeted groups (excluding one-year and two-year scholarship recipients).[14] Approximately 97% (682) of the 707 cases excluded from the imputed sample were public school non-applicants who were missing a

[12] The multiple imputation software package, NORM – developed and written by J.L. Schafer, Department of Statistics, Pennsylvania State University – was used to create imputed data sets for the analyses conducted and presented in this report. The NORM software is available as a free download at: http://www.stat.psu.edu/~jls/. For more information about Multiple Imputation, see: (a) Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. J. Wiley & Sons, New York; or (b) Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London.

[13] Although demographic and classroom characteristics were included in the imputation model as predictors of missing achievement scores, missing demographic and classroom data were not imputed. Therefore, analyses of student demographic and classroom characteristics include only cases that provided data; cases with missing data were dropped from these analyses.

[14] The sample size of 2,302 reported here resulted after (a) excluding 111 one-year scholarship recipients and 79 two-year scholarship recipients from the original sample of 2428, and (b) adding 64 misidentified applicant non-recipients to the sample.

considerable amount of achievement data across the measures and testing episodes. Nevertheless, the public school non-applicants included in the imputed data set still comprised the largest group of students (513). In addition, one case was excluded from the three-year scholarship recipient group; 14 cases were excluded from the former scholarship recipient group; and 10 cases were excluded from the scholarship winner non-user group due to a lack of achievement and demographic data. All of the four-year scholarship recipients and all of the applicant non-recipients were included in the imputed data set.

Table 4 below compares the number of students in the original sample to the number of students in the imputed sample by group. Table 4 also displays the number of cases in the imputed sample that were complete and the number of cases that were missing at least one achievement score prior to imputation. Table 5 displays the amount of missing data estimated and replaced on each test in the sample selected for imputation (N = 1,595). Finally, Figure 1 graphically displays the number of students in each group for whom a complete set of achievement data was imputed.[15]

Table 4. Sample Selected for Imputation vs. Original Sample

Scholarship Status	Original 2001 Sample	Imputed 2001 Sample		Missing Data in the Sample Used for Imputation	
	N	N	% of Original Sample	Complete Cases	Cases with missing Data
3 year Scholarship Recipients	120	**119**	99.2%	76	**43 (36%)**
4 year Scholarship Recipients	361	**361**	100.0%	235	**126 (35%)**
Former Scholarship Recipients	112	**98**	87.5%	29	**69 (70%)**
Scholarship Winner Non-users	95	**85**	89.5%	30	**55 (65%)**
Applicant Non-recipients	419	**419**	100.0%	179	**240 (57%)**
Non-applicants	1195	**513**	42.9%	326	**187 (37%)**
Total N =	2302[a]	**1595**	69.3%	875	**720 (45%)**

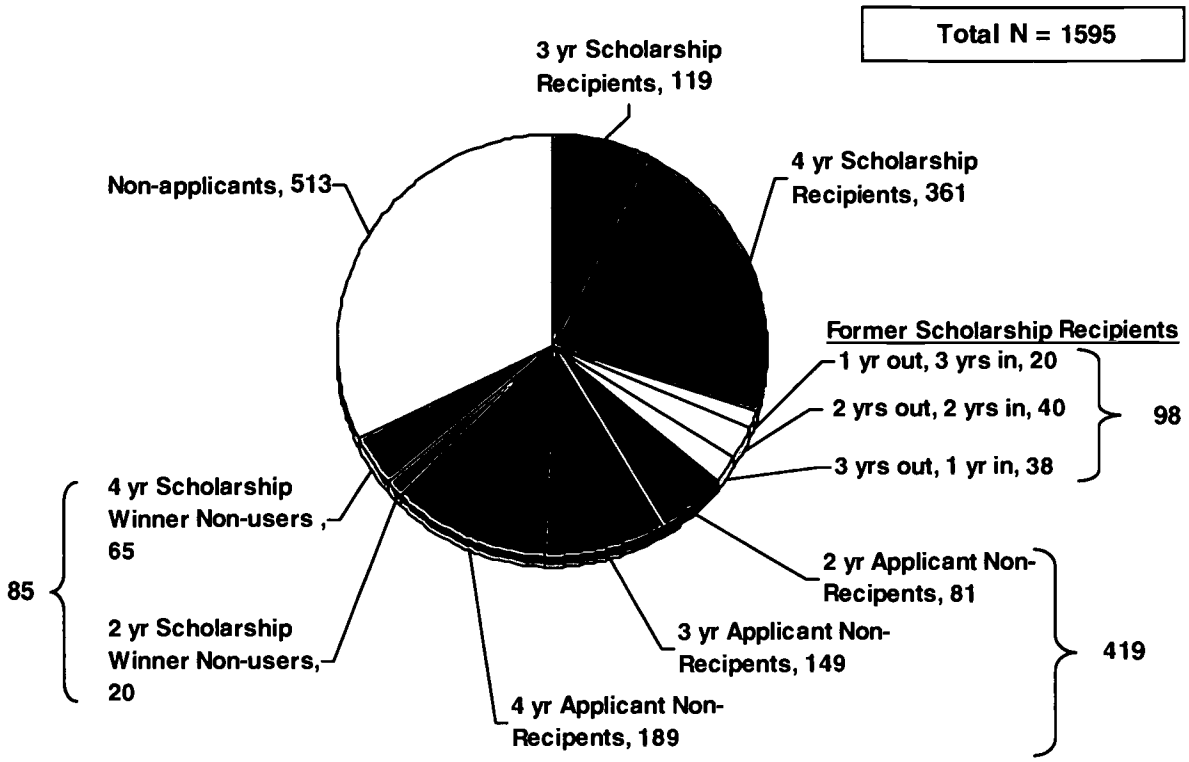
a. 2428 in original sample – (111 one-year scholarship recipients) – (79 two-year scholarship recipients) + (64 misidentified applicant non-recipients) = 2302

[15] For the analyses involving one-year and two-year scholarship recipients, a total of 79 two-year recipients who entered the program in second grade and 111 one-year recipients who entered the program in third grade were available – contingent upon the completeness of their data on a per-analysis basis.

Table 5. Sample Selected for Imputation: Amount of Missing Data Estimated and Replaced on Each Achievement Measure

Achievement Measure & Testing Episode	Valid – Observed Cases		Missing Data: Scores Imputed for 2001 Data Set		Total in 2001 Imputation Sample
	N	Percent	N	Percent	N
Reading					
Fall 1st Grade (1998)	1390	87.1%	205	**12.9%**	1595
Spring 1st Grade (1999)	1358	85.1%	237	**14.9%**	1595
Spring 2nd Grade (2000)	1345	84.3%	250	**15.7%**	1595
Spring 3rd Grade (2001)	1414	88.7%	181	**11.3%**	1595
Language					
Fall 1st Grade (1998)	1388	87.0%	207	**13.0%**	1595
Spring 1st Grade (1999)	1358	85.1%	237	**14.9%**	1595
Spring 2nd Grade (2000)	1348	84.5%	247	**15.5%**	1595
Spring 3rd Grade (2001)	1411	88.5%	184	**11.5%**	1595
Mathematics					
Fall 1st Grade (1998)	1394	87.4%	201	**12.6%**	1595
Spring 1st Grade (1999)	1380	86.5%	215	**13.5%**	1595
Spring 2nd Grade (2000)	1359	85.2%	236	**14.8%**	1595
Spring 3rd Grade (2001)	1409	88.3%	186	**11.7%**	1595
Total Achievement					
Fall 1st Grade (1998)	1354	84.9%	241	**15.1%**	1595
Spring 1st Grade (1999)	1310	82.1%	285	**17.9%**	1595
Spring 2nd Grade (2000)	1303	81.7%	292	**18.3%**	1595
Spring 3rd Grade (2001)	1351	84.7%	244	**15.3%**	1595

Figure 1. Imputed Sample used for 2001 Analyses



2.3 Data Analysis Techniques

Data analysis has been conducted to focus on each of the three evaluation questions guiding this phase of the evaluation and on relevant emerging sub-questions. For each question, analyses included both descriptive and inferential statistical techniques. Inferential analyses relied upon analysis of variance approaches, multiple regression analysis, and, when appropriate, follow-up pairwise comparisons.[16]

Overview of Data Analysis by Evaluation Question

Question 1, regarding the demographic characteristics of scholarship recipients versus the various comparison groups, has been investigated using descriptive approaches as well as univariate and multivariate analyses of variance to examine the data collected during the fourth testing episode (spring of the 2000-2001 school year). When appropriate, the Holm's

[16] Throughout the analysis and results section of this report, the F-statistics and p-values associated with follow-up pairwise comparisons are not reported in the text but, instead, are presented in Appendix B for interested readers.

Sequentially Rejective Dunn-Sidak procedure has been used to conduct post hoc (i.e., follow-up) comparisons to further examine indicated differences.[17] In each set of analyses, students serve as the unit of analysis and, because missing demographic data were not imputed, the largest possible sample sizes have been used for each analysis. Because the specific students included in each set of analyses are not exactly the same, some caution must be exercised when interpreting the findings pertaining to Question 1.

An additional component of Question 1 involves the comparison of students who entered the CSTP as kindergartners, first-graders, second-graders, and third graders (i.e., the question of whether and how the demographic characteristics of students differ as a function of when students entered the CSTP). Descriptive and inferential analyses have been conducted on student demographic and achievement data to address this question. In addition, for students who recently entered the CSTP in second and third grade, descriptive analyses identify both the schools in which these students were enrolled prior to receiving and using a scholarship (pubic or private) and the students' past involvement in the program as applicant non-recipients, winner non-users, or non-applicants. Furthermore, the demographic characteristics of recent scholarship recipients who previously attended public schools have been contrasted with the characteristics of recent recipients who attended private schools when these students were awarded scholarships. Analyses also have been conducted to investigate differences in demographic characteristics between students who remained in the program continuously and those who chose to withdraw from the CSTP and return to public school (i.e., examining the effect of differential exit from the program).

Question 2, regarding the classroom and teacher characteristics of private versus public schools, utilizes the largest possible sample for each of three independent sets of analyses. Unlike analyses of Question 1, which utilized students as the unit of analysis, Question 2 analyses relied on data aggregated by classroom. Moreover, because the primary comparison for this question is

[17] Holm's multiple comparison method has been selected because it offers greater statistical power (i.e., the ability to detect effects based on sample data if effects are truly present in the population of interest) than do other follow-up comparison procedures, while still controlling the Type I error rate across multiple comparisons at less than 0.05 (family-wise alpha). The Type I error rate is the probability of finding sufficient evidence to conclude that an effect (e.g., a mean difference between two groups) is present based on sample data when, in fact, *no effect* is present in the population of interest (under the statistical assumption that no effect/mean difference exists and sufficient evidence must be found to reject this a priori assumption).

between public and private schools, rather than between or among the groups of students, this analysis investigates differences in classroom and teacher characteristics by classroom and between public and private schools. Descriptive, multivariate, and univariate analyses of variance have been conducted to address this question and related sub-questions regarding the characteristics of the classrooms that students attended during the 2000-2001 school year.

Question 3, addressing the academic achievement of students in the various comparison (scholarship-status) groups, was investigated in a somewhat different manner than either of the two questions discussed above. In an effort to determine whether differential patterns of achievement change occurred across time as a function of the students' scholarship status, it was believed important to examine not only overall differences in achievement between the student groups, but also to focus attention on the pattern of achievement displayed by each group from the beginning of first grade through the end of third grade. Consequently, data analyses for Question 3 relied primarily on mixed-design analyses of covariance.[18] Using group membership as a between-subjects variable and time (testing episode) as a repeated-measures (within-subject) variable, these analyses provide the opportunity to investigate the impact of program participation across time while statistically controlling for initial differences among the comparison groups on key demographic characteristics. Given the nature of the data collected, a mixed-design analysis of covariance approach is believed to be the most defensible and easily interpreted.

Thus for Question 3, the analyses for each achievement measure (reading, language, mathematics, and total score) examined: (a) differences in academic achievement among the scholarship recipients, applicant non-recipients, non-applicants, and winner non-users, regardless of when they were tested; (b) differences in academic achievement among each of the four testing episodes, regardless of group membership; and, most importantly for the present evaluation, (c) the interaction of group membership and testing episode – to determine whether the achievement of scholarship students changed at a different rate across time in comparison to students in the public school groups. Together, these analyses address the issue of whether or not

[18] For a discussion of the use and interpretation mixed-design analysis of variance (in this case, one-between/one-within subjects independent variable), see J. Stevens, *Applied Multivariate Statistics for the Social Sciences, 3rd.* Edition, (Mahwah, NJ: Erlbaum Associates, 1996); J. J. Kennedy and A. J. Bush, *An Introduction to the Design and Analysis of Experiments in Behavioral Research*, (Lanham, MD: University Press of America, 1985).

participation in the CSTP has a beneficial effect on academic achievement in comparison to students who remain in public schools.

When appropriate, follow-up comparisons have been performed to explicate the nature of these effects. One set of follow-up comparisons was designed to examine differences in achievement among the comparison groups within each of the four testing episodes. Another set of comparisons was designed to examine differences among the comparison groups in the amount of achievement score change that occurred from one testing episode to another (i.e., across time). The latter set of comparisons was conducted to determine in what manner student achievement changed differentially among the target groups from kindergarten to third grade. Taken together, both sets of follow-up comparisons converge to describe the nature and magnitude of indicated differences among the groups.

Analyses of Question 3 also have been conducted to examine possible differences in achievement among scholarship students who entered the program as kindergartners, first graders, second graders, and third graders (i.e., based on differential entry into the program). Similarly, analyses were conducted to examine achievement differences between students who continuously remained in the CSTP and those students who chose to withdraw from the program and return to public school (i.e., examining the effect of differential exit from the program on academic achievement).

3.0 Analyses and Results

3.1 Question One

What are the characteristics of students who participate in the Cleveland Scholarship and Tutoring Program and how do they compare with students who do not participate?

Analyses of Question One include descriptions and comparisons of the demographic characteristics of students in each of the primary groups (i.e., scholarship winner users, winner non-users attending public schools, public school applicant non-recipients, and public school non-applicants), using the largest available sample for each. Univariate analyses were conducted for each of the following demographic characteristics: student sex, student minority status, estimated meal code as a proxy for family income, and student mobility (i.e. the number of school changes made since kindergarten, 1997-1998). The results of the analyses of sex, minority status, and estimated meal code are presented together below.[19] The findings from the analyses of student mobility are presented in a separate section.

3.1.1 Student Demographic Characteristics: Spring, 2001 (late third grade)

To examine whether scholarship recipients in the 2000-2001 school year possessed different demographic characteristics than their counterparts in the public schools, the demographic characteristics of all scholarship recipients, regardless of when they entered the Cleveland Scholarship and Tutoring Program (CSTP), were compared to the three primary public school groups (i.e., applicant non-recipients, non-applicants, and scholarship winner non-users). Table 6 presents descriptive statistics associated with the demographic characteristics of students in the four primary groups of interest that were targeted in the spring of third grade, 2001.

Separate univariate analyses were conducted on each of three demographic variables: sex, minority status, and estimated meal code. These analyses identified no statistically significant

[19] For the analyses presented throughout the report, minority status has been dichotomously coded as Non-minority = 1, Minority = 2. The Non-minority group is comprised entirely of Caucasian students. The Minority group is mainly comprised of African American students, but it also includes students identified as Hispanic and Multi-racial. Similarly, student sex has been dichotomously coded as Male = 1, Female = 2. Therefore, when group means are presented for minority status or sex, the numbers following the decimal indicate the percentage of students who are of minority status or female. (e.g., A mean of 1.54 for minority status indicates that 54% of students in that group have been identified as minority students).

differences in sex among the groups, ($F[3, 1611] = 1.29$, $p = .28$), but revealed statistically significant between-group differences in minority status ($F[3, 1668] = 16.42$, $p < .001$) and in estimated meal code ($F[3, 1505] = 22.39$, $p < .001$). Table 7 presents expanded minority status data for the groups, and Figure 2 graphically presents these data.

Table 6. Student Demographic Data: Spring 2001 (Late Third Grade)

Student group	Sex[a]			Minority Status[b]			Estimated Meal Code[c]		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Scholarship Recipients	**1.48**	.500	613	**1.66**	.474	667	**1.68**	.807	660
Public Applicant Non-Recipients	**1.54**	.499	406	**1.81**	.394	418	**1.41**	.717	325
Public Non-Applicants	**1.50**	.500	513	**1.77**	.422	503	**1.40**	.731	455
Public Winner Non-Users	**1.53**	.502	83	**1.92**	.278	84	**1.14**	.463	69
Total	**1.50**	.500	1615	**1.75**	.434	1672	**1.51**	.769	1509

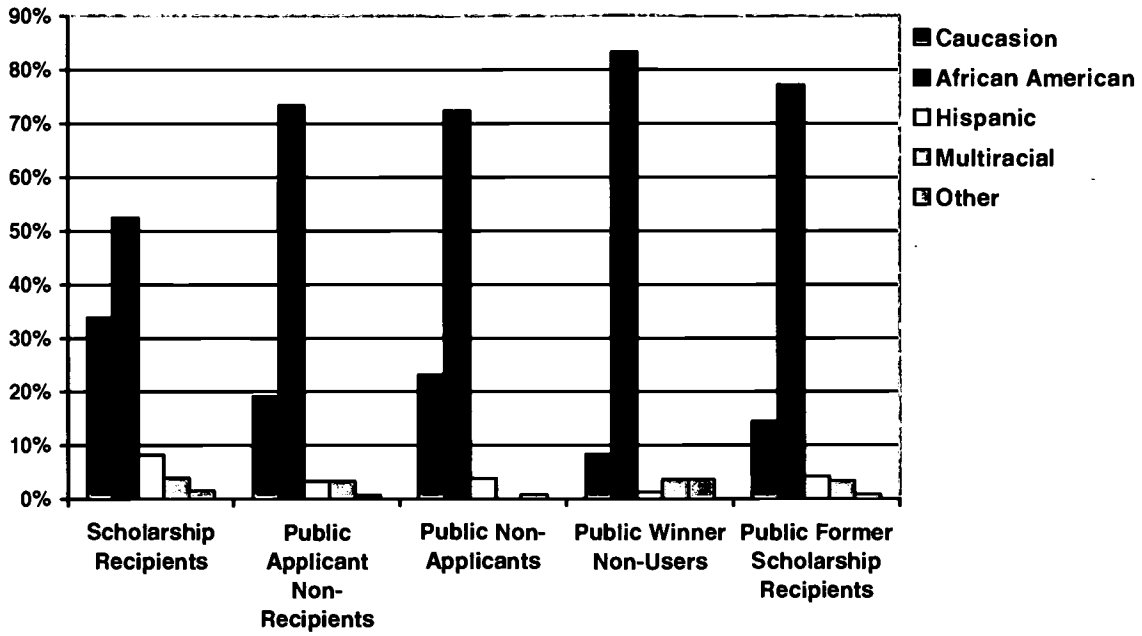
a. Sex was coded dichotomously as Male = 1, Female = 2. As a result, numbers following the decimal indicate the percentage of students who were female. This interpretation can be applied to all subsequently reported analyses of minority status
b. Minority Status was coded dichotomously as Non-minority = 1, Minority = 2. As a result, numbers following the decimal indicate the percentage of students who were of minority status. This interpretation can be applied to all subsequently reported analyses of minority status.
c. Estimated meal code was coded as 1 = eligible for free lunch, 2 = eligible for reduced lunch, 3 = eligible for neither free nor reduced lunch (i.e., required to pay).

Table 7. Expanded Minority Status Data: Spring 2001 (Late Third Grade)

		Caucasian	African American	Hispanic	Multiracial	Other	Total
Scholarship Recipients	**Percent**	**33.8%**	**52.5%**	**8.2%**	**3.9%**	**1.5%**	**100%**
	N	226	351	55	26	10	668
Public Applicant Non-Recipients	**Percent**	**19.1%**	**73.4%**	**3.3%**	**3.3%**	**0.7%**	**100%**
	N	80	307	14	14	3	418
Public Non-Applicants	**Percent**	**23.1%**	**72.4%**	**3.8%**	**0.0%**	**0.8%**	**100%**
	N	116	364	19	0	4	503
Public Winner Non-Users	**Percent**	**8.3%**	**83.3%**	**1.2%**	**3.6%**	**3.6%**	**100%**
	N	7	70	1	3	3	84
Total	**Percent**	**25.6%**	**65.3%**	**5.3%**	**2.6%**	**1.2%**	**100%**
	N	429	1092	89	43	20	1673

Racial Composition of the Cleveland Municipal School District (average daily membership = 76,323): 19.7% Caucasian, 70.6% African American, 8% Hispanic, 0.5% Multiracial, and 1.2% Other. These data were obtained from the Cleveland Municipal School District's 1999-2000 Annual Report, which is available for public download at http://www.cmsdnet.net/administration/publications.htm

Figure 2. 2000-2001 Minority Status by Category and Scholarship Group



Follow-up pairwise comparisons[20] of minority status data indicate that applicant non-recipients (81% minority), non-applicants (77% minority), and scholarship winner non-users (92% minority) each were comprised of significantly more minority students than was the scholarship recipient group (66% minority). In other words, by 2000-2001, fewer minority students (i.e., African-American, Hispanic, or Multiracial students) were in the scholarship recipient group than were in any of the public school comparison groups. There were, however, no statistically significant differences in minority status among public school non-applicants, public school applicant non-recipients, and public school scholarship winner non-users.

Although the overall proportion of minority students in the public school groups (79% minority) was greater than the proportion in the scholarship recipient group (66% minority), this difference primarily reflects the disproportionate representation of African Americans in the public school groups (73.7%) versus the scholarship recipient group (52.5%). However, the other racial and ethnic groups classified as minority students in the analyses described above (i.e., Hispanic and

[20] All pairwise comparisons throughout this report were conducted using the Holm's Sequentially Rejective Dunn-Sidak procedure to adjust for conducting multiple pairwise comparisons and control overall experimentwise error at $\alpha < .05$. For information on this procedure, see: Kirk, R.E. (1995). Experimental design: Procedures for the behavioral science (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Multiracial students) descriptively comprised a greater proportion of the scholarship recipient group in comparison to the public school groups as a whole (see Table 7). Specifically, during the 2000-2001 school year, the proportion of scholarship recipients who were Hispanic (8.2%) was statistically greater than the proportion of public school students who were Hispanic (3.4%).[21] Although, the proportion of scholarship recipients in 2000-2001 who were classified as Multiracial (3.9%) was descriptively greater than the proportion of Multiracial students represented in the public school comparison groups (1.7%), the difference was not statistically significant. In other words, the observed difference in multiracial student composition cannot be meaningfully interpreted.

Pairwise comparisons of meal code status[22] found that the average meal code estimate for scholarship recipients (M = 1.68) was significantly higher than the average meal code estimate for each of the public school comparison groups: applicant non-recipients (M = 1.41), non-applicants (M = 1.40), and scholarship winner non-users (M = 1.14). That is, regardless of the duration of participation in the CSTP, scholarship recipients in 2000-2001 had the highest average estimated meal code of the four groups contrasted in the analysis. Furthermore, the average estimated meal code both for the public applicant non-recipients (M = 1.41) and for the public non-applicants (M = 1.40) was statistically higher than the estimated meal code for the public winner non-users (M = 1.14). The difference in average estimated meal code between public school applicant non-recipients and public school non-applicants, however, was not statistically significant.

Overall in 2000-2001, a significantly greater proportion of public school students were African-American in comparison to students whose families were awarded and chose to use a scholarship. However, a greater proportion of scholarship recipients were Hispanic in comparison to the public school groups as a whole. Furthermore, students who were awarded a

[21] The total number of Hispanic students in the data set is 89. A nonparametric binomial test indicated that the percentage distribution of the 89 Hispanic students who were scholarship recipients (62%) and public school students (38%) differed at statistically significant level from an hypothesized even 50/50 division between the groups (p = .034). That is, the greater proportion of Hispanic students observed in the scholarship recipient group as opposed to the public school group likely reflects a systematic difference that cannot be attributed to chance alone.

[22] Estimated meal codes are used as a proxy measure of family income throughout this report, such that higher meal codes roughly correspond to higher family incomes.

scholarship but chose to enroll in public schools (i.e., winner non-users) were of substantially lower income than were students in any other group, whereas scholarship users tended to be from families of higher income in comparison to the other groups. In addition, each group consisted nearly equally of males and females.

3.1.2 Student Demographic Characteristics and Academic Achievement

Regression analyses were performed to evaluate whether student demographic characteristics were related to academic achievement. Reading, language, mathematics, science, social studies, and total scale scores from the spring, 2001 testing episode were separately regressed on student sex, minority status, and estimated meal code as a proxy for family income.[23] Collectively, the set of student demographic variables accounted for a statistically significant, but small, portion of the variance in *reading* ($F[3, 1506] = 32.12$, $R^2 = .060$, $p < .001$); *language* ($F[3, 1506] = 37.58$, $R^2 = .070$, $p < .001$); *mathematics* ($F[3, 1503] = 29.99$, $R^2 = .056$, $p < .001$); *total* ($F[3, 1502] = 41.27$, $R^2 = .076$, $p < .001$); *science* ($F[3, 1377] = 35.93$, $R^2 = .073$, $p < .001$); and *social studies* scores ($F[3, 1377] = 27.14$, $R^2 = .056$, $p < .001$). Each individual demographic variable accounted for a statistically significant unique portion of the variance in 2000-2001 reading, language, mathematics, total achievement, and social studies scores. Only student minority status and estimated meal code, however, explained a statistically significant portion of the variance in 2000-2001 science achievement scores. Table 8 displays the results of this analysis by achievement measure and demographic characteristic.

Notably, the nature of the relationship between each demographic characteristic and academic achievement was consistent across the achievement measures. Female students were associated with higher performance than were male students, non-minority students were associated with higher performance than were minority students, and high meal codes were associated with higher performance than were low meal codes.[24] Although these relationships were found to be

[23] Previous levels of the Terra Nova administered to students during the first grade and second grade testing episodes did not include a social studies or science section. The social studies and science achievement measures were first administered to third graders in spring, 2001 (Terra Nova Level 13). Furthermore, the total scale score represents the average of a student's reading, language, and mathematics scale scores, regardless of the test-level administered.

[24] Correlations among the demographic characteristics found a statistically significant relationship between minority status and estimated meal code. Minority students were associated with lower meal codes (i.e., lower family income) than were non-minority students. No other correlations among gender, minority status, and estimated meal codes were statistically significant.

statistically significant, the amount of variance in 2000-2001 achievement scores accounted for by the demographic characteristics was small (i.e., less than 8% across measures).

Table 8. 2000-2001 Achievement Score Regression Analysis Results: Achievement Scores Regressed on Student Demographic Characteristics

Demographic Characteristic	Reading				Language				Mathematics			
	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p
Sex	**10.53**	.135	.018	< .001	**10.33**	.146	.021	< .001	**7.33**	.095	.009	< .001
Minority Status	**-11.91**	-.133	.017	< .001	**-12.67**	-.156	.023	< .001	**-15.88**	-.179	.031	< .001
Meal Code	**6.60**	.130	.016	< .001	**5.73**	.125	.015	< .001	**4.57**	.091	.008	< .001

Demographic Characteristic	Total				Science				Social Studies			
	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p
Sex	**9.45**	.141	.020	< .001	**4.55**	.049	.002	.059	**6.06**	.078	.006	.003
Minority Status	**-13.29**	-.172	.028	< .001	**-24.70**	-.233	.052	< .001	**-14.69**	-.167	.026	< .001
Meal Code	**5.67**	.130	.016	< .001	**5.19**	.086	.007	.001	**5.89**	.118	.013	< .001

Note: **Beta** coefficients indicate the magnitude and direction of the linear relationship between each demographic characteristic and the corresponding achievement measure. Specifically, the unstandardized betas indicate the amount of change in the predicted scale score per one-unit of change in the predictor variable. The standardized beta coefficients are presented to facilitate meaningful comparisons among the demographic predictors.

The sr^2 columns display the squared *semi-partial correlations* between each demographic characteristic and the corresponding criterion achievement measure. The squared semi-partial correlation indicates the amount of unique variance in achievement scores that a given demographic predictor explains, after statistically accounting for any variance in achievement that is shared (i.e., overlaps) with the other two demographic predictors in the regression model.

Additional regression analyses were performed to determine whether student demographic characteristics were related to academic achievement *change* across the four year period. Change-scores representing the algebraic difference in scale scores between the first testing episode (early first grade, 1998) and the most recent testing episode (late third grade, 2001) were created for each student as an index of change in academic achievement from first grade to third grade. Reading, language, mathematics, and total changes-scores were separately regressed on student sex, minority status, and estimated meal.[25] Collectively, the set of student demographic variables accounted for a small but statistically significant portion of the variance in achievement change across three years in *reading* ($F[3, 1470] = 3.71$, $R^2 = .008$, $p = .011$); *language* (F[3,

[25] Science and social studies achievement measures were first administered in third grade, 2001. Therefore, these measures could not be included in the change-score analyses.

1470] = 2.73, R^2 = .006, p = .043); *mathematics* (F[3, 1470] = 4.42, R^2 = .009, p = .004); and *total score* (F[3, 1470] = 6.45, R^2 = .013, p < .001).

Student sex emerged as the best independent predictor of achievement change from first grade to third grade, after controlling for the other demographic characteristics. Specifically, female students were associated with greater amounts of positive achievement score change (gain) in each of the four subtest areas in comparison to male students. Table 9 displays the results of this analysis by achievement measure and demographic characteristics. It is important to emphasize that the amount of variance in achievement change across time explained by demographic characteristics – both independently and collectively – was rather small (less than 1% on each test). In particular, the demographic characteristics collectively accounted for less variance in achievement score *change across the four testing episodes* than they did in achievement *within each testing episode* (see the findings above).

Table 9. 1998 to 2001 Achievement Change Score Regression Analysis Results: Achievement Change Scores Regressed on Student Demographic Characteristics

Demographic Characteristic	Reading Change (1st – 3rd Grade)				Language Change (1st – 3rd Grade)				Mathematics Change (1st - 3rd Grade)			
	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	P	Beta	Std. Beta	sr^2	p
Sex	**5.31**	.067	.005	.010	**4.12**	.055	.003	.034	**6.03**	.082	.007	.002
Minority Status	**5.08**	.055	.003	.039	**4.34**	.049	.002	.063	**3.71**	.043	.002	.104
Meal Code	**.54**	.011	< .001	.692	**- .05**	- .001	< .001	.969	**- .48**	- .010	< .001	.709

Demographic Characteristic	Total Change (1st – 3rd Grade)			
	Beta	Std. Beta	sr^2	p
Sex	**5.21**	.092	.008	< .001
Minority Status	**4.48**	.067	.004	.011
Meal Code	**0.03**	.001	< .001	.978

Note: **Beta** coefficients indicate the magnitude and direction of the linear relationship between each demographic characteristic and the corresponding achievement measure. Specifically, the unstandardized betas indicate the amount of change in the predicted scale score per one-unit of change in the predictor variable. The standardized beta coefficients are presented to facilitate meaningful comparisons among the demographic predictors.

The **sr^2** columns display the *squared semi-partial correlations* between each demographic characteristic and the corresponding criterion achievement measure. The squared semi-partial correlation indicates the amount of unique variance in achievement scores that a given demographic predictor explains, after statistically accounting for any variance in achievement that is shared (i.e., overlaps) with the other two demographic predictors in the regression model.

3.1.3 Student Mobility

Student mobility was calculated by tracking the number of times that each student changed schools from kindergarten (1997-1998) to third grade (2000-2001). Therefore, the student mobility indicator ranges from zero (no school changes/ one school attended from grade K-3) to three (the student changed schools every year from grade K-3). Because several students in the sample attended one school, moved to a different school, and then changed back to attend their original school (i.e., they made two school changes but only attended two different schools), the *number of school changes* made by students from kindergarten to third grade was used in the following analyses of student mobility as opposed to the total number of schools attended during this time period. For the most part, however, two school changes indicate that a student attended three different schools. In contrast, zero school changes always indicate that one school was attended. For each target group, Table 10 displays the number and proportion of students in each mobility category as well as the average number of school changes made by each group from kindergarten to third grade. For comparison purposes, non-applicants from private schools have been included in this analysis.[26]

A univariate analysis of variance revealed that the mean number of school changes made from kindergarten to third grade differed at a statistically significant level between at least two of the student groups ($F[5, 1688] = 5.50$, $p < .001$). Follow-up paired comparisons found that all of the groups, except three-year scholarship recipients, changed schools a greater number of times from kindergarten through third grade than did private school non-applicants. In other words, private school non-applicants demonstrated more stability (less mobility) in their school enrollment ($M = .07$) than did four-year scholarship recipients ($M = .23$), public applicant non-recipients ($M = .21$), public non-applicants ($M = .25$), and public scholarship winner non-users ($M = .24$). Three-year scholarship recipients ($M = .17$) did not differ significantly from private school non-applicants in the number of school changes made, and they also did not differ significantly from

[26] Former scholarship recipients and recent entrants into the CSTP (i.e., one-year and two-year scholarship recipients) are not included in Table 10. Students in the former scholarship recipient groups, on average, attended two schools over the period covered in this investigation, and only one or two students in each group attended three schools. This is not surprising because former scholarship recipients, by definition, attended private schools for one or more years and then withdrew from the program to attend public schools. Students who recently entered the CSTP (in second and third grade) were excluded from the mobility analyses because (1) school of enrollment data prior to entrance into the program were unavailable for many students, and (2) almost all of the students for whom school enrollment data were available attended two schools from kindergarten through third grade, as expected.

the other groups in the analysis. No statistically significant differences in the average number of school changes were found among four-year scholarship recipients, public applicant non-recipients, public non-applicants, and public scholarship winner non-users.

Most notably, the four-year scholarship recipients made a greater number of school changes (M = .23) than did their peers in private schools whose families never applied for a scholarship (M = .07). This means that four-year scholarship recipients, on average, changed schools slightly more frequently within the private school system than did private school non-applicants – who predominantly attended the same school across all four years (K-3). However, the mobility of four-year scholarship recipients, as measured by the number of school changes that they made, was not statistically different from the mobility of the public school comparison groups in the analysis.

Table 10. Student Mobility: The Number and Relative Frequency of Students in Each Mobility Category

Student Group	Number of School Changes made from Kindergarten (1997-1998) through Third Grade (2000-2001)								Mean # of School Changes (K-3)	
	0		1		2		3			
	N	%	N	%	N	%	N	%	Mean	SD
3-year Scholarship Recipients (N=119)	101	**84.9%**	17	**14.3%**	0	–	1	**0.8%**	**.17**	0.43
4-year Scholarship Recipients (N=361)	286	**79.2%**	68	**18.8%**	7	**1.9%**	0	–	**.23**	0.46
Applicant Non-recipients (N=419)	337	**80.4%**	74	**17.7%**	8	**1.9%**	0	–	**.21**	0.45
Winner Non-users (N=85)	67	**78.8%**	16	**18.8%**	2	**2.4%**	0	–	**.24**	0.47
Non-applicants: Public (N=512)	392	**76.6%**	111	**21.7%**	8	**1.6%**	1	**0.2%**	**.25**	0.48
Non-applicants: Private (N=198)	185	**93.4%**	13	**6.6%**	0	–	0	–	**.07**	0.24
Total: N = 1694	1368	**80.8%**	286	**16.9%**	25	**1.5%**	2	**0.1%**		

Student Mobility as a Function of Minority Status, Sex, and Meal Code

Separate univariate analyses of variance were conducted to investigate whether the number of school changes that were made from kindergarten to third grade differed between minority and non-minority students; between male and female students; or among free, reduced, and paid lunch students (as a proxy for low, moderate, and high income families, respectively). Factorial

analyses of variance were conducted to determine not only whether mobility differed as a function of these student demographic characteristics, but also whether any differences in mobility based on minority status, gender, or estimates meal codes varied as a function of student group membership (i.e., to determine whether any student group by demographic characteristic interactions were present). Because demographic data were not obtained from private school non-applicants, this comparison group was excluded from the following analyses. Table 11 displays the mean number of school changes made from kindergarten to third grade as a function both of minority status and of group membership.

Table 11. Mean Number of Schools Changes (K-3) as a Function of Minority Status and Scholarship Status

	Minority Status						Marginal Means Scholarship Status		
	Minority			Non-Minority					
Student Group	Mean	SE	N	Mean	SE	N	Mean	SE	N
3-year Scholarship Recipients	.20	.05	82	.11	.08	37	.15	.05	119
4-year Scholarship Recipients	.29	.03	240	.11	.04	121	.20	.03	361
Public Applicant Non-Recipients	.21	.03	338	.24	.05	80	.22	.03	418
Public Non-Applicants	.28	.02	387	.19	.04	116	.23	.03	503
Public Winner Non-Users	.26	.05	77	.00	.18	7	.13	.09	84
Marginal Means Minority Status	.25	.01	1124	.13	.04	361			

The analysis of mobility (number of school changes made from grades K-3) as a function of minority status and student group revealed a statistically significant main effect of minority status ($F[1, 1475] = 6.77$, $p = .009$), but neither the main effect of student group membership nor the student group by minority status interaction was statistically significant. The statistically non-significant main effect of student group indicates that, regardless of minority status, none of the groups differed significantly in the number of school changes made from kindergarten through third grade. This was also found in the previous analysis that included private school non-applicants.

The statistically significant main effect of minority status, however, indicates that minority students made a statistically greater number of school changes, on average, across the four

school years under consideration (M = .25) than did non-minority students (M = .13), regardless of their group status. In addition, the statistically non-significant minority status by student group interaction indicates that the association between minority status and mobility (i.e., minority students changing schools more frequently than non-minority students) did not differ depending on the group to which students belong.

Similar to the previous analysis of minority status and student group membership, the analysis of mobility as a function of sex and student group membership found a statistically significant main effect of sex ($F[1, 1442] = 5.39$, $p = .020$), but neither the main effect of student group membership nor the student group by sex interaction was statistically significant. The statistically non-significant main effect of student group membership indicates that, regardless of sex, none of the student groups differed significantly in the number of school changes made from kindergarten through third grade. The statistically significant main effect of sex, however, indicates that male students made a statistically greater number of school changes, on average, across the four school years under consideration (M = .26) than did female students (M = .18), regardless of their student group status. In addition, the statistically non-significant sex by group membership interaction indicates that the association between sex and mobility (i.e., male students changing schools more frequently than female students) did not differ as a function of group membership. Table 12 displays the mean number of schools changes that were made from kindergarten to third grade as a function both of sex and of student group.

Table 12. Mean Number of School Changes (K-3) as a Function of Sex and Scholarship Status

	Gender						Marginal Means Scholarship Status		
	Male			Female					
Student Group	Mean	SE	N	Mean	SE	N	Mean	SE	N
3-year Scholarship Recipients	**.22**	.06	58	**.09**	.06	54	**.16**	.04	112
4-year Scholarship Recipients	**.21**	.03	158	**.22**	.03	181	**.22**	.03	339
Public Applicant Non-Recipients	**.27**	.03	185	**.18**	.03	221	**.22**	.02	406
Public Non-Applicants	**.27**	.02	257	**.24**	.02	255	**.25**	.02	512
Public Winner Non-Users	**.31**	.07	39	**.18**	.07	44	**.25**	.05	83
Marginal Means Minority Status	**.26**	.02	697	**.18**	.02	755			

An analysis of mobility as a function of estimated meal code and student group membership, however, failed to reveal any statistically significant effects of these factors on student mobility. That is, the main effect of meal code, the main effect of scholarship status, and the meal code by scholarship status interaction all were statistically non-significant with respect to student mobility.

Because the differences in mobility both between males and females and between minority students and non-minority students did not differ depending on the group to which the students belonged, student group membership was excluded from the following analysis, and minority status and sex both were entered as factors in a univariate analysis of variance. As was the case in the previous analyses, the main effect both of minority status and of sex was statistically significant, such that: (a) minority students changed schools more frequently than did non-minority students, regardless of sex, and (b) male students changed schools more frequently than did female students, regardless of minority status.

An inspection of the means displayed in Table 13 reveals that, descriptively, male minority students made the greatest number of school changes from kindergarten to third grade (M = .28), followed in descending order by: female minority students (M = .23), male non-minority students (M = .19), and finally female non-minority students (M = .13) who made the fewest number of school changes. The sex by minority status interaction, however, was not statistically significant. That is, the association between minority status and mobility was statistically the same for females as it was for males; and conversely, the association between gender and mobility was the same for minority students as it was for non-minority students. Therefore, the observed differences described above and displayed in Table 13 were not statistically significant and cannot be meaningfully interpreted.

Table 13. Mean Number of School Changes (K-3) as a Function of Sex and Minority Status

	Gender						Marginal Means Minority Status		
	Male			Female					
Minority Status	Mean	SE	N	Mean	SE	N	Mean	SE	N
Minority	**.28**	.02	512	**.23**	.01	574	**.25**	.01	1086
Non-Minority	**.19**	.03	175	**.13**	.03	180	**.16**	.03	355
Marginal Means Gender	**.24**	.02	687	**.18**	.02	754			

Student Mobility and Academic Achievement

In terms of the relationship between student mobility and academic achievement, the number of school changes that students made between kindergarten and third grade largely was unrelated to the measures of achievement obtained over the course of this evaluation. Student mobility demonstrated a small, but statistically significant, inverse relationship with only the language achievement and total achievement measures at the end of second grade (language, $r = -.08$, $p = .002$; total, $r = -.06$, $p = .013$) and at the end of third grade (language, $r = -.06$, $p = .023$; total, $r = -.05$, $p = .044$). This finding indicates that students who changed schools frequently tended to achieve at slightly lower levels on the language achievement measure, and by extension on the total achievement measure, at the end of both second grade and third grade than did students who remained at the same school. Student mobility, however, was not associated with the other achievement measures (reading and mathematics) at any point from kindergarten to third grade. Moreover, student mobility was unrelated to the amount of change in achievement demonstrated by students between any of the testing episodes, regardless of the achievement measure analyzed.[27]

3.1.4 Demographic Characteristics and Differential Entry of Scholarship Students

Embedded within the question of whether and how the demographic characteristics of scholarship recipients differ from the public school comparison groups are two relevant subquestions. First, are students who enter the CSTP later (e.g., as second or third graders as opposed to kindergartners or first graders) different from those students who enter earlier? Second, are the demographic characteristics of scholarship recipients who withdraw from the program at various points from kindergarten through third grade different from scholarship participants who remain in the program? The first subquestion, pertaining to differential entry into the program and student demographic characteristics, is discussed below. Differential exit from the CSTP is discussed in Section 4.1.7.

[27] It should also be noted that, despite the statistically significant effects reported in this section, student mobility largely was invariant across the four school years under consideration. That is, most of the students attended the same school from grade K-3, such that none of the average mobility indices were greater than one. Therefore, the practical significance of these findings may be limited.

To address the first subquestion regarding differential entry, univariate analyses of variance were conducted contrasting the 2000-2001 (students' third grade year) demographic characteristics of students who entered the program as: (a) kindergartners in autumn, 1997, (b) first graders in autumn, 1998, (c) second graders in autumn, 1999, and (d) third graders in autumn, 2000. Table 14 presents descriptive statistics for the demographic characteristics of students based on differential entry into the CSTP and, for comparison purposes, the demographic characteristics of the public school groups.

Table 14. Student Demographic Data, Late Third Grade (2001): Differential Entry

Student Group	Sex			Minority Status			Estimated Meal Code		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
1 year Scholarship Recipients	**1.37**	.486	94	**1.60**	.492	108	**2.03**	.862	110
2 year Scholarship Recipients	**1.40**	.493	68	**1.68**	.468	79	**1.49**	.716	78
3 year Scholarship Recipients	**1.48**	.502	112	**1.69**	.465	119	**1.60**	.777	117
4 year Scholarship Recipients	**1.53**	.500	339	**1.66**	.473	361	**1.65**	.790	355
Average across Scholarship Recipient Groups	**1.48**	.500	613	**1.66**	.474	667	**1.68**	.807	660
Public Applicant Non-Recipients	**1.54**	.499	406	**1.81**	.394	418	**1.41**	.717	325
Public Non-Applicants	**1.50**	.500	513	**1.77**	.422	503	**1.40**	.731	455
Public Winner Non-Users	**1.53**	.502	83	**1.92**	.278	84	**1.14**	.463	69

Statistically significant differences in student sex ($F[3, 609] = 3.41$, $p = .017$) and meal code status ($F[3, 656] = 9.20$, $p < .001$) were found among scholarship recipients who entered the program at various points over the past four years. However, no statistically significant differences in minority status were found among these groups of scholarship recipients ($F[3, 663] = .76$, $p = .51$). Approximately 66% of scholarship students in each year of entry were minority.

Follow-up pairwise comparisons indicate that scholarship recipients who entered the program as kindergartners in autumn, 1997, and have remained in the program for four years, were comprised of significantly more female students (53% female) than were scholarship recipients who entered the program as third graders in autumn, 2000 (37% female). No other statistically

significant differences in sex were found among the groups of scholarship recipients who entered the program at different points from kindergarten to third grade.

Pairwise comparisons of estimated meal code revealed that students who entered the CSTP as third graders in autumn, 2000 had significantly higher meal codes (M = 2.03) than did students who entered the CSTP as: kindergartners (M = 1.65); first graders (M = 1.60); and second graders (M = 1.49). No statistically significant differences in estimated meal code status, however, were found among the groups of scholarship recipients who entered the program prior to third grade.

In addition, the four groups of scholarship recipients, identified based on the year in which they entered the program, each were contrasted with the public school comparison groups to further explicate the demographic similarities and differences between scholarship students and public school students.

Entry in Kindergarten, 1997: Univariate analyses comparing the 2000-2001 demographic characteristics of the public school groups to those of scholarship recipients who entered the program as kindergartners in autumn, 1997 and have remained in the program (i.e., four-year scholarship recipients) revealed between group differences in minority status ($F[3, 1362] = 11.98$, $p < .001$) and estimated meal code ($F[3, 1200] = 13.65$, $p < .001$). However, no statistically significant between-group differences in student sex emerged from this analysis. Follow-up comparisons of minority status found that four-year scholarship recipients were comprised of significantly fewer minority students (66% minority) than were applicant non-recipients (81% minority), non-applicants (77% minority), or winner non-users (92% minority). In addition, comparisons of estimated meal codes found that four-year scholarship recipients had a significantly higher average meal code (M = 1.65) than did any of the public school comparison groups. The estimated meal code for four-year scholarship recipients was significantly higher than that of applicant non-recipients (M = 1.41), non-applicants (M = 1.40), and winner non-users (M = 1.14).

Entry in First Grade, 1998: Univariate analyses comparing the 2000-2001 demographic characteristics of the public school groups to those of scholarship recipients who entered the program as first graders in autumn, 1998 and have remained in the program (i.e., three-year

scholarship recipients) revealed between-group differences in minority status ($F[3, 1120] = 5.83$, $p = .001$) and estimated meal code ($F[3, 962] = 5.90$, $p = .001$). However, no statistically significant between-group differences in student sex emerged from this analysis. As was the case with four-year scholarship recipients, pairwise comparisons of minority status found that three-year scholarship recipients were comprised of significantly fewer minority students (69% minority) than were either applicant non-recipients (81% minority) or winner non-users (92% minority). The difference in minority status between three-year scholarship recipients and non-applicants (77% minority), however, was not statistically significant. Follow-up comparisons of estimated meal codes found that three-year scholarship recipients had a significantly higher mean meal code ($M = 1.60$) than any of the public school comparison groups ($M = 1.41$, 1.40, and 1.14 for applicant non-recipients, non-applicants, and winner non-users, respectively).

Entry in Second Grade, 1999: Univariate analyses that included scholarship recipients who entered the program as second graders in autumn, 1999 (i.e., two-year scholarship recipients) and the public school comparison groups revealed statistically significant between-group differences in minority status ($F[3 , 1080] = 5.28$, $p = .001$) and estimated meal code ($F[3, 923] = 3.36$, $p = .018$), but no statistically significant between-group differences in student sex emerged from this analysis. As was the case with three-year recipients, follow-up comparisons revealed that two-year scholarship recipients were comprised of significantly fewer minority students (68% minority) than were either applicant non-recipients (81% minority) or winner non-users (92% minority), and the difference between two-year scholarship recipients and non-applicants (77% minority) was not statistically significant. Additionally, comparisons revealed that two-year scholarship recipients had a significantly higher average estimated meal code ($M = 1.49$) than did winner non-users ($M = 1.14$). No statistically significant differences in estimated meal code were found between two-year scholarship recipients and either applicant non-recipients ($M = 1.41$) or non-applicants ($M = 1.40$).

Entry in Third Grade, 2000: Univariate analyses that included scholarship recipients who entered the program as third graders in autumn, 2000 (i.e., one-year scholarship recipients) and the public school comparison groups revealed statistically significant between-group differences in sex ($F[3, 1092] = 3.17$, $p = .024$), minority status ($F[3, 1109] = 10.70$, $p < .001$), and estimated meal code ($F[3, 955] = 28.28$, $p < .001$). Follow-up comparisons revealed that statistically fewer

female students belonged to the one-year scholarship recipient group (37% female) than belonged to the public school applicant non-recipient group (54% female). No statistically significant differences in student sex, however, were found between one-year scholarship recipients and either applicant non-recipients or winner non-users. As was the case with four-year recipients, comparisons revealed that one-year scholarship recipients were comprised of significantly fewer minority students (60% minority) than were applicant non-recipients (81% minority), non-applicants (77% minority), and winner non-users (92% minority). The average estimated meal code for one-year scholarship recipients (M = 2.03) also was significantly higher than the average estimated meal code for applicant non-recipients (M = 1.41), non-applicants (M = 1.40), and winner-non-users (M = 1.14).

Summary: Differential Entry

Since the fall of 1997, the demographic characteristics of the cohort of scholarship students now in third grade have changed. Although the proportion of minority scholarship students has not changed significantly over time (approximately 66% minority in each group of new entrants from 1997-1998 to 2000-2001), the proportion of male students who entered the CSTP as third graders in 2000-2001 is significantly greater than the proportion of male students who entered the program as kindergarteners in 1997-1998. Furthermore, recent third-grade entrants tend to be from families with significantly higher incomes in comparison to students who entered the CSTP as kindergartners in 1997, first graders in 1998, or second graders in 1999.

In comparison to the public school groups, the scholarship recipient groups tend to have fewer minority students. However, only the most recent entrants into the CSTP (in third grade, 2000-2001) and the initial entrants (in kindergarten, 1997-1998) have a significantly smaller proportion of minority students in comparison to all of the public school groups. The proportion of minority students who entered the program as first graders (in 1998-1999) and second graders (in 1999-200) does not differ significantly from the proportion of minority non-applicants in public schools. First and second grade entrants differ from only the public school applicant non-recipients and scholarship winner non-users, in that they are comprised of a greater proportion of minority students.

Furthermore, with the exception of scholarship recipients who entered the program in second grade (1999-2000), the average estimated meal code (family income) of scholarship recipients is significantly greater than that of all three public school comparison groups. The family income of second grade entrants into the program, however, is statistically similar to the family income of public school applicant non-recipients and non-applicants, but greater than that of winner non-users.

3.1.5 Students' Status Prior to Becoming Scholarship Recipients

An issue of interest in the evaluation of voucher use is whether or not participants would have attended private schools even without the voucher. To address this question, data on school of prior enrollment (public or private) and prior scholarship status (non-applicant, applicant non-recipient, or winner non-user) were analyzed for students who entered the CSTP as third graders in autumn, 2000. Two sets of analyses were conducted of prior school of enrollment and prior scholarship status for (a) students who entered the CSTP as second graders in autumn, 1999 and for (b) those students who entered as third graders in autumn, 2000.

Data were available from 76 scholarship recipients who entered the program as second graders, and from 110 scholarship recipients who entered the program as third graders. Table 15 displays the proportional breakdown of prior school-type (public vs. private) and scholarship status (non-applicant, applicant non-recipient, or scholarship winner non-user) for these students.

Table 15. Previous Scholarship Status and School Type of Students Entering the Program in Second Grade (autumn, 1999)

Prior School Type in 1st Grade (1998-1999)	Prior Scholarship Status in 1st Grade (1998-1999)						Prior School Type Totals	
	Non-applicant		Applicant Non-recipient		Winner Non-user			
	N	% of total N	N	% of total N	N	% of total N	N	% of total N
Public	**6**	7.9%	**6**	7.9%	**1**	1.3%	**13**	17.1%
Private	**21**	27.6%	**8**	10.5%	**0**	–	**29**	38.2%
Unknown	**31**	40.8%	**3**	4.0%	**0**	–	**34**	44.7%
Prior Scholarship Status Totals	**58**	76.3%	**17**	22.4%	**1**	1.3%	Total N = **76**	

Of the 76 scholarship recipients who entered the program in second grade, most had never applied for a scholarship in the past (76.3%), and, of those whose former school of enrollment was known, a majority (69.0%) had attended a private school as a first grader. Of the 29 students who entered the program as second graders from *private schools*, eight had applied for but not received a scholarship during first grade (27.6%) and the remaining 21 students (72.4%) were non-applicants during that year. Among the 13 students who entered the CSTP as second graders after attending a *public school* in first grade, six (46.2%) had applied for but not been awarded a scholarship; six were non-applicants (46.2%); and one (7.7%) had been awarded a scholarship but had chosen not to use it in first grade. Of the 34 scholarship recipients who entered the program as second graders after attending an unknown type of school in first grade, three students (8.8%) were applicant non-recipients and 31 (91.2%) were non-applicants in first grade.

Table 16 displays the proportional breakdown of prior school-type (public vs. private) and scholarship status (applicant non-recipient, non-applicant, or scholarship winner non-user) for students who entered the program as third graders in the 2000-2001 school year. Based on data collected during the first and second grade testing episodes, the schools of prior enrollment are not known for 41 of the 110 students who entered the CSTP in third grade. Of the 69 remaining students, 47 (68.1%) had attended a *private school* during first grade, second grade, or both. In contrast, 22 students (31.9%) entered the program in third grade after attending a *public school* during first or second grade.

With regard to the students' scholarship status prior to entering the program in third grade, 18 students (16.4% of 110) were applicant non-recipients from private schools both in first and second grade, whereas five students (4.5% of 110) were applicant non-recipients from public schools both in first and second grade. Only seven private school students (6.4% of 110) and three public school students (2.7% of 110) were non-applicants both in first grade and in second grade. Furthermore, only four students (3.6%) who entered the program as third graders were scholarship winner non-users in previous years.

Table 16. Previous Scholarship Status and School Type of Students Entering the Program in Third Grade (autumn, 2000)

Prior School Type in 1st and 2nd Grade		Scholarship Status in 1st Grade (98-99) & in 2nd Grade (99-00)					
		Non App 1st & 2nd	ANR 1st & 2nd	WNU 1st & 2nd	Non App 1st ANR 2nd	Non App 1st WNU 2nd	School Type Totals
Former PUBLIC School	Public 1st & Public 2nd	3	5	1	4	–	13 (11.8%)
	Public in 1st & Unknown in 2nd	1	1	–	–	–	2 (1.8%)
	Unknown in 1st & Public 2nd	1	–	–	4	2	7 (6.4%)
	Subtotal: Public %	**5** (4.6%)	**6** (5.5%)	**1** (0.9%)	**8** (7.3%)	**2** (1.8%)	**22** (20.0%)
Former PRIVATE School	Private 1st & Private 2nd	7	18	1	3	–	29 (26.4%)
	Private 1st & Unknown in 2nd	1	1	–	1	–	3 (2.7%)
	Unknown in 1st & Private 2nd	5	5	–	5	–	15 (13.6%)
	Subtotal: Private %	**13** (11.8%)	**24** (21.8%)	**1** (0.9%)	**9** (8.2%)	–	**47** (42.7%)
Unknown in 1st & Unknown in 2nd		**38** (34.5%)	**1** (0.9%)	–	**2** (1.8%)	–	**41** (37.3%)
Scholarship Status Totals		**56** (50.9%)	**31** (28.2%)	**2** (1.8%)	**19** (17.3%)	**2** (1.8%)	Total N = **110**

Note: Scholarship Recipient Group abbreviations: **Non App** = *Non-applicants*, **ANR** = *Applicant Non-recipients*, **WNU** = *Scholarship Winner Non-users*

Summary: Student Status Prior to Becoming Scholarship Recipients

Overall, proportionally more students who entered the CSTP in second and third grade, combined, were drawn from private schools (68% of 111) as opposed to public schools (32% of

111), and the difference was statistically significant.[28] However, the proportion of second and third grade entrants into the CSTP who had applied for and been denied a scholarship in preceding years (58% of 106) was statistically equivalent to the proportion of second and third grade entrants who had never previously applied for a scholarship (42% of 106). Therefore, although recent entrants tend to come from private as opposed to public schools, overall they are equally as likely to have been non-applicants as they are to have been applicant non-recipients prior to entering the program.

However, separate analyses based on the specific year in which students entered the CSTP found that the distribution of students between (a) previous non-applicants and (b) previous applicant non-recipients differed depending on whether students entered in second grade or third grade. Students who attended private schools prior to entering the program in third grade were statistically more likely to have applied for and been denied a scholarship in preceding years than they were to have never previously applied for a scholarship. Similarly, former public school students who entered as third graders were statistically more likely to have been applicant non-recipients as opposed to non-applicants in one or more of the preceding years.[29] Therefore, regardless of their prior school of enrollment, third-grade entrants were more likely to have been applicant non-recipients as opposed to non-applicants in one or more of the preceding years.[30]

In contrast to students who entered the program in third grade, private school recipients who entered in second grade were statistically more likely to have never previously applied for a

[28] Among all students with a known school of prior enrollment (N=111), a non-parametric binomial test found that the percentage of students who entered the CSTP from *private schools* (68%) versus *public schools* (32%) differed at a statistically significant level from an hypothesized even 50%/50% division ($p < .0001$). The observed breakdown between former private and public school students also differs significantly from any hypothesized (i.e., expected) division in which the percentage of new scholarship recipients from public schools is expected to be greater than the percentage from private schools.

[29] Among students who entered the CSTP in third grade (2000-2001) from *private schools*, the distribution between prior applicant non-recipients (33, or 72% of 46) and prior non-applicants (13, or 28% of 46) differed at a statistically significant level from an hypothesized even 50%/50% division ($p = .005$). Similarly, among students who entered the CSTP in third grade (2000-2001) from *public schools*, the distribution between prior applicant non-recipients (14, or 74% of 19) and prior non-applicants (5, or 26% of 19) differed at a statistically significant level from an hypothesized even 50%/50% division ($p = .05$).

[30] Among only third grade (2000-2001) entrants into the CSTP, a Chi-square test of independence found that (a) prior school of enrollment was unrelated to (b) prior scholarship status ($\chi 2[1] = .025$, $p > .05$). That is, the differential breakdown of third-grade entrants between previous non-applicants and previous applicant non-recipients does not vary depending on whether a student entered from a public versus private school.

scholarship than they were to have applied for and been denied a scholarship the year before.[31] Moreover, unlike third-grade entrants, public school students who entered the program in second grade were equally distributed between previous non-applicants and previous applicant non-recipients. As was the case with third grade entrants, however, prior school of enrollment was statistically unrelated prior scholarship status for second grade entrants into the CSTP.[32]

3.1.6 Demographic Characteristics of Recent Scholarship Recipients from Public versus Private Schools

A question of interest to the current investigation is whether the demographic characteristics of students who enter the CSTP from public schools differ significantly from the characteristics of students already enrolled in private schools when they receive a scholarship. As the analyses presented above indicate, recent scholarship recipients as a whole (i.e., all students who entered the program in second and third grade) are *less likely* to be minority students and *more likely* to be from families with higher incomes than either (a) recipients who entered the program in kindergarten and first grade or (b) their peers in public schools. This trend, however, was not found when analyses of minority status and family income (estimated meal codes) are separately conducted on students who attended public versus private schools prior to entering the program.

Minority Status

Among recent scholarship recipients with known schools of prior enrollment, a Chi-square test for independence revealed that *school of prior enrollment* (public versus private) was related to *minority status* (minority versus non-minority). This means that the proportional distribution of minority and non-minority students who entered the CSTP from public schools in second and third grade differed significantly from the distribution of minority and non-minority students who entered from private schools during the same time period ($\chi^2[1] = 10.40$, $p = .001$). Table 17 displays the frequency of recent minority and non-minority entrants into the program both from private and from public schools of prior enrollment. Descriptively, a greater proportion of

[31] Among students who entered the CSTP in second grade (1999-2000) from *private schools*, the distribution between prior non-applicants (21, or 72% of 29) and applicant non-recipients (8, or 28% of 46) differed at a statistically significant level from an hypothesized even 50%/50% division ($p = .026$).

[32] Among only second grade (1999-2000) entrants into the CSTP, a Chi-square test of independence found that (a) prior school of enrollment was unrelated to (b) prior scholarship status ($\chi 2[1] = 1.896$, $p = .17$). That is, the breakdown of second-grade entrants between previous non-applicants and previous applicant non-recipients does not vary at a level beyond chance depending on whether a student entered from a public versus private school.

entrants from public schools were minorities (78%) as opposed to non-minorities (22%), whereas entrants from private schools were approximately evenly divided between minorities (46%) and non-minorities (54%).

Table 17. Frequency and Proportion of Second-grade and Third-grade Entrants into the CSTP by Minority Status and School of Prior Enrollment

School of Prior Enrollment		Minority Status		
		Minority	Non Minority	Row Totals
Public	N	28	8	36
	% of all Public	78%	22%	100%
	% of Minority Status group	43%	15%	31%
	% of Total	24%	7%	31%
Private	N	37	44	81
	% of all Private	46%	54%	100%
	% of Minority Status group	57%	85%	69%
	% of Total	32%	38%	69%
Column Totals	N	65	52	Total N = 117
	% across Pub./Priv. groups	56%	44%	
	% of Minority Status group	100%	100%	
	% of Total	56%	44%	

Of particular interest is whether scholarship recipients who enter the program after attending public schools in previous years possess demographic characteristics similar to their former classmates in public schools. In other words, are students from public schools who use a scholarship to attend private schools similar to or different from their peers who remain in public schools? In the present sample of third grade students, 79% of 995 public school students are minorities and 21% are non-minorities (i.e., Caucasians),[33] and these proportions are statistically equivalent to the proportions of minority (78%) and non-minority (22%) students who previously attended *public schools* and entered the CSTP in second and third grade. In contrast, the proportions of minority (46%) and non-minority (54%) students who previously attended *private schools* and entered the CSTP in second and third grade are statistically different from the

[33] The proportions of minority and non-minority public school students in the sample is statistically equivalent to the population proportions of minority (80%) and non-minority (20%) students in the Cleveland Municipal School District, K-12 (These data were obtained from the Cleveland Municipal School District's 1999-2000 Annual Report, which is available for public download at http://www.cmsdnet.net/administration/publications.htm)

proportions of minority and non-minority students in public schools.[34] Therefore, in terms of minority status, students in the present sample from public schools who use a scholarship to attend private schools are similar to their peers who remain in public schools (i.e., 78-80% minority students), whereas scholarship recipients from private schools are significantly more likely to be non-minority (Caucasian) students.

Estimated Meal Codes (Family Income)

Similarly, among recent scholarship recipients with known schools of prior enrollment, a Chi-square test for independence revealed that *school of prior enrollment* (public versus private) was related to *estimated meal code status* (free and reduced lunches versus paid lunches). This means that the proportional distribution of lower income students who entered the CSTP from public schools in second and third grade differed significantly from the distribution of lower income students who entered from private schools during the same time period ($\chi^2[1] = 11.33$, $p = .001$). Table 18 displays the frequency of recent lower income entrants into the program (who qualified for free or reduced price lunches) and higher income entrant (who did not qualify for free or reduced lunched) both from private and from public schools of prior enrollment. Descriptively, a greater proportion of scholarship entrants from public schools were from lower income families (87%) as opposed to higher income families (13%), whereas entrants from private schools were approximately evenly divided between lower income (54%) and higher income (46%) families.

Of particular interest is whether scholarship recipients who enter the program after attending public schools in previous years possess demographic characteristics similar to their former classmates in public schools. In other words, are students from public schools who use a scholarship to attend private schools similar to or different from their peers who remain in public schools? In the present sample of third grade students, 87% of 927 public school students are from lower income families (i.e., they receive or qualify for free or reduced lunches) and 13% are from higher income families (i.e., they do not qualify for free or reduced lunches. These proportions are statistically equivalent to the proportions of lower income (87%) and higher

[34] Non-parametric binomial tests, compared the observed proportion of minority students who entered the CSTP in second grade (1999-00) and third grade (2000-01) with the observed proportion of minority students in public schools. The binomial tests revealed: (1) The proportion of minority scholarship recipients from *public schools* (78%) did not differ significantly from the public school proportion (79% minority, $p = .51$); but (2) The proportion of minority scholarship recipients from *private schools* (46%) differed significantly from the public school proportion (79% minority, $p < .001$)

income (13%) students who previously attended *public schools* and entered the CSTP in second and third grade. In contrast, the proportions of lower income (54%) and higher income (46%) students who previously attended *private schools* and entered the CSTP in second and third grade are statistically different from the public school proportions.[35] Therefore, in terms of estimated meal code status (family income) , students in the present sample from public schools who use a scholarship to attend private schools are similar to their peers who remain in public schools (i.e., 87% lower income), whereas scholarship recipients from private schools are significantly more likely to be from families with higher incomes.

Table 18. Frequency and Proportion of Second-grade and Third-grade Entrants into the CSTP by Estimated Meal Code Status and School of Prior Enrollment

School of Prior Enrollment		Meal Code Status: Free or Reduced (lower income)	Meal Code Status: Paid (higher income)	Row Totals
Public	N	32	5	37
	% of all Public	87%	13%	100%
	% of Meal Code group	43%	12%	32%
	% of Total	28%	4%	32%
Private	N	43	36	79
	% of all Private	54%	46%	100%
	% of Meal Code group	57%	88%	68%
	% of Total	37%	31%	68%
Column Totals	N	75	41	Total N = 116
	% across Pub./Priv. groups	65%	35%	
	% of Meal Code group	100%	100%	
	% of Total	65%	35%	

Summary: Demographic Characteristics of Recent Scholarship Recipients from Public versus Private Schools

With respect to the proportion of minority students and students from families with lower income, recent scholarship recipients, who entered the CSTP in second grade (1999-00) and third grade (2000-01) after attending public schools in previous years program, were statistically

[35] Non-parametric binomial tests, compared the observed proportion of lower income students who entered the CSTP in second grade (1999-00) and third grade (2000-01) with the observed proportion of lower income students in public schools. The binomial tests revealed: (1) The proportion of lower income scholarship recipients from *public schools* (87%) did not differ significantly from the public school proportion (87% lower income, $p = .56$); but (2) The proportion of lower income scholarship recipients from *private schools* (54%) differed significantly from the public school proportion (79% minority, $p < .001$)

equivalent to their classmates who remained in public schools – approximately 80% minority and 87% lower income. In contrast, scholarship recipients who entered the program in second grade and third grade after attending private schools in previous years differed significantly from their peers in public schools. Recent recipients who attended *private schools* prior to entering the CSTP were approximately 46% minority and 54% lower income.

Furthermore, among the 72 recent scholarship recipients whose previous schools of enrollment are unknown, both the proportion of minority students (76%) and the proportion of lower income students (85%) were statistically equivalent to the proportion of minority (78%) and lower income (87%) students in the public schools. Although the present data set does not identify their schools of enrollment prior to entering the program, CSTP office records suggest that many of these students attended public schools before being awarded scholarships. Table 19 below displays the expanded demographic characteristics of recent scholarship recipients from public schools contrasted with the characteristics of recent recipients from private and unknown schools of prior enrollment.

Table 19. Expanded Student Demographic Data, Late Third Grade (2001): Differential Entry and Prior School of Enrollment

Scholarship Recipient Group	Prior School of Enrollment	Minority Status					Estimated Meal Code (Income Proxy)			
		Caucasian	African American	Hispanic	Other	N	Free (lower)	Reduced	Paid (higher)	N
1 year Scholarship Recipients (3rd grade entry)	Public	27%	50%	23%	–	22	48%	35%	17%	23
	Private	61%	31%	4%	4%	49	13%	23%	65%	47
	Unknown	21%	59%	15%	5%	39	55%	25%	20%	40
	Total	40%	45%	12%	4%	110	35%	26%	39%	110
2 year Scholarship Recipients (2nd grade entry)	Public	14%	78%	7%	–	14	78%	14%	7%	14
	Private	44%	41%	6%	9%	32	47%	34%	19%	32
	Unknown	27%	67%	3%	3%	33	75%	16%	9%	32
	Total	32%	58%	5%	5%	79	64%	23%	13%	78

Racial Composition of the Cleveland Municipal School District (average daily membership = 76,323): 19.7% Caucasian, 70.6% African American, 8% Hispanic, 1.7% Other (Asian, Native American, Multi-ethnic). These data were obtained from the Cleveland Municipal School District's 1999-2000 Annual Report, which is available for public download at http://www.cmsdnet.net/administration/publications.htm

3.1.7 Demographic Characteristics and Differential Exit of Scholarship Recipients

A second demographic subquestion involves the issue of differential exit from the Cleveland Scholarship and Tutoring Program. In other words, are the demographic characteristics of students who leave the CSTP after one or more years of participation different from the characteristics of those students who remain in the program? To address this question, three groups of former scholarship recipients have been identified. Each group is composed of students who withdrew from the program and returned to the public school system after one to three years of participation.

Specifically, as of the 2000-2001 school year, the following groups of former scholarship recipients have been identified: (a) students who participated in the program as kindergartners during the 1997-1998 school year, withdrew from the program after kindergarten, and subsequently have spent three years attending public schools from first through third grade; (b) students who participated in the program both as kindergartners and as first graders, withdrew from the program after first grade (the 1998-1999 school year), and subsequently have spent two years attending public schools from second through third grade; and (c) students who participated in the program from kindergarten through second grade, withdrew from the program after second grade (the 1999-2000 school year), and have spent one year attending public schools during third grade (the 2000-2001 school year). Based on the number of years that former scholarship recipients have attended public schools since exiting the program, these groups are referred to as three-year, two-year, and one-year former recipients, respectively.

To examine the demographic characteristics associated with differential exit from the CSTP, univariate analyses were conducted comparing the 2000-2001 demographic characteristics of students who have received and continuously used scholarships since kindergarten in 1997 (i.e., four-year scholarship recipients) with the aforementioned groups of former scholarship recipients. Table 20 presents descriptive statistics for the demographic characteristics associated with these groups.[36]

[36] Meal code data could be obtained for only five former scholarship recipients who withdrew from the program after three years of participation (i.e., one-year former scholarship recipients). As a result, the analysis of estimated meal code data did not include one-year former recipients.

Table 20. Student Demographic Data, Late Third Grade (2001): Differential Exit

Student group	Sex			Minority Status			Estimated Meal Code		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
1 year Former Scholarship Recipients	1.47	.513	19	1.75	.444	20	--	--	--
2 year Former Scholarship Recipients	1.45	.504	38	1.92	.267	40	1.27	.521	30
3 year Former Scholarship Recipients	1.53	.506	36	1.87	.343	38	1.38	.711	24
4 year Scholarship Recipients	1.53	.500	339	1.66	.473	361	1.65	.790	355
Public Applicant Non-Recipients	1.54	.499	406	1.81	.394	418	1.41	.717	325
Public Non-Applicants	1.50	.500	513	1.77	.422	503	1.40	.731	455
Public Winner Non-Users	1.53	.502	83	1.92	.278	84	1.14	.463	69

Univariate analyses revealed statistically significant differences both in minority status ($F[3, 455] = 5.93$, $p = .001$) and in estimated meal code ($F[2, 406] = 4.52$, $p = .011$) among the former scholarship recipient groups, but no statistically significant differences in sex were found ($F[3, 428] = .41$, $p = .75$). Follow-up pairwise comparisons were conducted to examine the differences in minority status between four-year scholarship recipients and former recipients. The comparisons indicate that a statistically smaller proportion of scholarship recipients who remained in the program from kindergarten through third grade are of minority status (66%) than are either three-year former recipients who withdrew from the program after one year of participation (87%) or two-year former recipients who withdrew from the program after first grade (92%). Three-year former recipients, however, did not differ in minority status from two-year former recipients. No other statistically significant differences in minority status emerged from the follow-up analyses based on differential exit from the CSTP.

Follow-up pairwise comparisons of estimated meal code indicate that four-year scholarship recipients (students who have participated in the program continuously since kindergarten) had a statistically higher average estimated meal code ($M = 1.65$) than did two-year former scholarship

recipients who returned to public schools after first grade (M = 1.27).[37] No other statistically significant differences in estimated meal code were found.

In addition, univariate analyses of the 2000-2001 demographic data were conducted to determine whether demographic differences exist between the former scholarship recipients and the primary public school comparison groups (i.e., applicant non-recipients, non-applicants, and scholarship winner non-users). Table 20, above, displays the descriptive statistics for the demographic characteristics associated both with the former scholarship recipient groups and with the public school comparison groups. No statistically significant differences, however, were found among these groups on any of the demographic variables in the analysis.

Summary: Differential Exit

Regardless of when former scholarship recipients withdrew from the program, they tended to be comprised of a greater number of minority students and students from families with lower incomes in comparison to scholarship recipients who continuously participated in the program from kindergarten through third grade. However, the demographic characteristics of former scholarship recipients do not differ significantly from their peers in public schools (applicant non-recipients, winner non-users, or non-applicants). That is, students who chose to exit the CSTP tend to be more demographically similar to public school students than they are to the scholarship recipients who have chosen to remain in the program through third grade, 2001.

[37] Estimated meal codes are used as a proxy measure of family income throughout this report, such that higher meal codes roughly correspond to higher family incomes.

3.2 Question Two

What are the characteristics of the classrooms and teachers to which scholarship students are exposed in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?

Unless otherwise noted, data for Question Two were analyzed using classrooms as the unit of analysis rather than students. That is, classroom and teacher characteristics were aggregated (averaged) across all of the students in each classroom who provided relevant data. Comparisons were made between public and private schools using classroom and teacher data collected during the 2000-2001 school year from 251 public school classrooms and 68 private school classrooms. Five variables of interest were included in the analyses discussed below: class size, teacher's highest degree, teacher's years of experience, teacher's years of service at his or her present school, and certification (i.e., whether or not a teacher is certified).

3.2.1 Classroom Characteristics of Public vs. Private Schools: Spring, 2001 (late third grade)

Table 21 presents descriptive statistics on classroom and teacher characteristics for the public and private schools attended by third grade students in the sample during the 2000-2001 school year. Multivariate analysis of variance, indicate significant differences between public and private school classrooms (Pillai's Trace $F[5, 297] = 3.10$, $p = .010$) suggesting the appropriateness of follow-up univariate analyses. Univariate analyses indicated no significant differences between the public and private schools in class size ($F[1, 301] = .04$, $p = .85$) and in the highest degree earned by the classroom teachers ($F[1, 301] = .45$, $p = .50$). Public and private school teachers did, however, differ significantly in: (a) years of overall teaching experience ($F[1, 301] = 5.68$, $p = .018$); (b) years of teaching experience at the present school ($F[1, 301] = 3.89$, $p = .05$), and (c) teaching certification ($F[1, 301] = 5.24$, $p = .023$). Private school teachers in this sample had 3.5 more years of experience than their public school counterparts ($M = 14.4$ and 10.9, respectively). Similarly, the average tenure for private school teachers at their present schools ($M = 7.3$) was approximately two years greater than was the average tenure for public school teacher at their present schools ($M = 5.4$). However, a statistically higher proportion of public school teachers were certified (96%) in comparison to private school teachers (88%).

Table 21. 2000-2001 Classroom Characteristics: Public Schools vs. Private Schools

Classroom & Teacher Characteristics	Public Schools (N = 243)		Private Schools (N = 60)	
	Mean	SD	Mean	SD
Class Size (# of Students)	**21.9**	5.50	**21.8**	5.57
Years of Teaching Experience - Overall	**10.9**	10.08	**14.4**	11.44
Years of Teaching Experience at Present School	**5.4**	6.25	**7.3**	8.02
Highest Degree Earned	**3.11**	1.09	**3.00**	1.25
Certification	**1.96**	0.20	**1.88**	0.32

Highest Degree Earned was coded as: 1 = No degree, 2 = BA/BS, 3 = BA/BS +, 4 = MA/MS, 5 = MA/MS +, 6 = Ed.S., 7 = ABD, 8 = Ph.D. *Certification* was coded as: 1 = No certification, 2 = Certification

3.2.2 Classroom Characteristics and Academic Achievement

The 2000-2001 teacher and classroom variables explained only a small proportion of the variance in student academic achievement (averaged at the classroom-level). Using 2000-2001 scale scores from the six available subtests, the five classroom and teacher variables collectively accounted for approximately 2% of the variance both in reading ($R^2 = .017$) and in language achievement ($R^2 = .019$); 3% of the variance in science achievement ($R^2 = .033$) and total score ($R^2 = .029$); and approximately 4% of the variance both in mathematics ($R^2 = .042$) and in social studies achievement ($R^2 = .036$). However, the set of demographic characteristics accounted for a statistically significant portion of variance in only the mathematics achievement scores ($F[5, 298] = 2.63$, $R^2 = .042$, $p = .024$). Table 22 displays the results of these regression analyses.

Although the classroom and teacher variables explained very little variance in academic achievement, class size emerged as the single best predictor of student achievement across the subtests. That is, after controlling for the other classroom and teacher characteristics, class size was found to be a statistically significant, albeit weak, independent predictor of academic achievement, such that *larger* class sizes were associated with higher scale scores on the achievement measures. Specifically, class size independently accounted for a unique, and statistically significant, portion of the variance in 2000-2001 scores in: reading (unique $R^2 = .016$, $t[298] = 2.17$, $p = .031$); language (unique $R^2 = .017$, $t[298] = 2.26$, $p = .024$); mathematics (unique $R^2 = .031$, $t[298] = 3.09$, $p = .002$); total (unique $R^2 = .024$, $t[298] = 2.77$, $p = .006$);

science (unique R^2 = .019, t[292] = 2.38, p = .018); and social studies (unique R^2 = .024, t[292] = 2.70, p = .007). Thus, for every additional student in a classroom, classroom average achievement scores increase by the following amounts: reading, 0.64 scale score units; language, 0.60 scale score units; mathematics, 0.96 scale score units; total achievement, 0.73 scale score units; science, 0.98 scale score units; and social studies, 0.81 scale score units.

Table 22. 2000-2001 Achievement Score Regression Analysis Results: Achievement Scores (Aggregated by Classroom) Regressed on Classroom and Teacher Characteristics

Classroom Characteristic	Reading				Language				Mathematics			
	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p
Class Size	**.64**	.127	.016	.031	**.60**	.132	.017	.024	**.96**	.178	.031	.002
Years of Experience (Overall)	**- .11**	- .043	.001	.596	**- .05**	- .019	< .001	.816	**- .03**	- .009	< .001	.907
Years at Present School	**.20**	.048	.001	.542	**.17**	.044	.001	.571	**.45**	.101	.006	.190
Highest Degree Earned	**.16**	.007	< .001	.915	**.04**	.002	< .001	.976	**-1.51**	- .057	.003	.356
Certification	**3.95**	.033	.001	.572	**3.08**	.028	.001	.626	**.53**	.004	< .001	.943

Classroom Characteristic	Total				Science				Social Studies			
	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p	Beta	Std. Beta	sr^2	p
Class Size	**.73**	.161	.025	.006	**.98**	.139	.019	.018	**.81**	.158	.024	.007
Years of Experience (Overall)	**- .06**	- .025	< .001	.753	**.07**	.018	< .001	.822	**- .03**	- .012	< .001	.877
Years at Present School	**.27**	.072	.003	.354	**.53**	.091	.004	.246	**.45**	.106	.006	.174
Highest Degree Earned	**-.43**	- .019	< .001	.755	**- .77**	- .023	< .001	.719	**- 1.45**	- .059	.003	.349
Certification	**2.61**	.024	.001	.679	**- 1.68**	- .010	< .001	.860	**.98**	.008	< .001	.886

Note: Beta coefficients indicate the magnitude and direction of the linear relationship between each demographic characteristic and the corresponding achievement measure. Specifically, the unstandardized betas indicate the amount of change in the predicted scale score per one-unit of change in the predictor variable. The standardized beta coefficients are presented to facilitate meaningful comparisons among the demographic predictors. The sr^2 columns display the squared semi-partial correlations between each demographic characteristic and the corresponding criterion achievement measure. The squared semi-partial correlation indicates the amount of unique variance in achievement scores that a given demographic predictor explains, after statistically accounting for any variance in achievement that is shared (i.e., overlaps) with the other two demographic predictors in the regression model.

Additional regression analyses examined the relation between classroom characteristics and measures of individual achievement *change* across the four testing episodes. For this analysis, however, students were used as the unit of analysis and new classroom and teacher variables were created representing the average class size, average years of teaching experience, average

years of experience at the present school, and average highest degree earned associated with the classrooms and teachers that each student was exposed to over the four data-collection episodes. To examine whether and how classroom and teacher characteristics were associated with changes in student achievement across time, achievement change (gain) scores were created, representing the algebraic difference in scale scores between the first testing episode (early first grade, 1998) and the most recent testing episode (late third grade, 2001). The achievement change scores separately were regressed on the set of average classroom and teacher variables described above. Furthermore, separate analyses were performed for the reading, language, mathematics, and total score components. Because social studies and science were included only for data collected in third grade, they have not been used in the present analyses.

The regression analyses revealed that, collectively, the classroom and teacher variables failed to account for a statistically significant portion of the variance in achievement change on any of the subtest scores. Specifically, for each measure of achievement (reading, language, math, and total), the classroom variables collectively explained less than 2% of the variance in achievement change from the first testing episode in early first grade to the most recent testing episode in late third grade. That is, the type of classroom that students experienced, on average across the four time periods, was not related to change in academic achievement. Moreover, none of the classroom variables independently accounted for a substantial unique portion of the variance in achievement change on any of the measures of achievement.

3.2.3 Differences in Classroom and Teacher Characteristics within Public and Private Schools

Independent univariate analyses of variance were used to examine whether and how teacher and classroom variables differed within public and private schools as a function of (a) minority status and (b) estimated meal code.[38] Specifically, the following analyses were conducted to examine whether students of particular minority status or family income *within a particular type of school* (e.g., within public or within private schools) attend classrooms with characteristics that differ from the classrooms attended by other students in the same type of school. For instance, are minority scholarship recipients in private schools experiencing a different type of classroom

[38] Student sex was not considered because no statistically significant differences in sex were found among the groups of interest: all 2000-2001 scholarship recipients, public school applicant non-recipients, and public school non-applicants. For more information on this finding, refer to results for Question 1 above.

environment than are non-minority scholarship recipients in private schools and, if so, are similar differences present in public schools? These analyses address the issue of whether certain groups of scholarship recipients (e.g., minority scholarship recipients), either intentionally or unintentionally, select or are assigned to schools that possess different characteristics than the schools selected by or assigned to other identifiable groups of students in the present sample.[39] All 2000-2001 scholarship recipients were included in the analysis, regardless of when they entered the program.

Class Size as a Function of Minority Status[40] and Student Group

A 2 (minority status: minority vs. non-minority) by 2 (student group: scholarship recipients vs. public school students)[41] between-subjects factorial analysis of variance was conducted using 2000-2001 class size as the dependent variable. Figure 3 displays class size as a function both of minority status and of student group membership. The analysis revealed a statistically significant main effect of minority status ($F[1, 1641] = 33.31$, $p < .001$). Minority students attended significantly smaller classes ($M = 22.6$ students) than did non-minority students ($M = 24.3$ students), regardless of whether they were scholarship recipients in private schools or public school students. The main effect of student group membership also was statistically significant ($F[1, 1641] = 4.65$, $p = .031$). Scholarship recipients, on average, attended slightly larger classes ($M = 23.8$) than did public school students ($M = 23.1$), regardless of whether they were minority or non-minority students. From a practical standpoint, however, this is a small mean difference in class size.

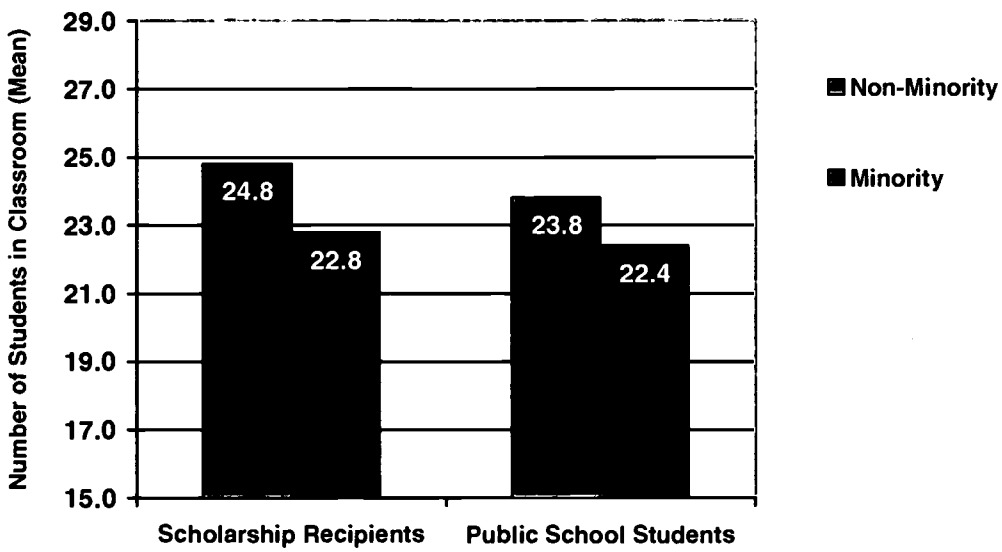
[39] The classroom characteristics experienced by distinct groups of students within each classroom are of interest. Therefore, individual students were used as the unit of analysis. Disaggregating the classroom and teacher data, however, has the effect of weighting each variable (e.g., years of teaching experience) by the number of students in each classroom. Consequently, data from classrooms with a large number of students received more weight in the following analyses than data from classrooms with few students. Furthermore, with the lager sample size resulting from disaggregating the data, small between-group differences could be found to be statistically significant. However, to determine the type of classroom environment that various groups of students experienced, it was necessary to use students as the unit of analysis.

[40] Fewer that 15% of minority students in each group were Hispanic, Multiracial, or designated as "other." Therefore, the following analyses effectively are comparing African-American students to Caucasian students.

[41] The aggregate public school comparison group includes four (4) subgroups of public school students: applicant non-recipients, non-applicants, winner non-users, and former scholarship recipients. This undifferentiated, aggregate public school comparison group is used in all of the analyses in this section; differences in teacher and classroom characteristics among the four constituent public school subgroups are discussed in the preceding section.

The student group by minority status interaction, however, was not statistically significant. Therefore, the difference in class size between minority and non-minority students did not vary as a function of student group membership. Both in public and in private schools (among scholarship recipients), minority students attend slightly smaller classes than did their non-minority peers.

Figure 3. 2000-2001 Mean Class Size as a Function of Minority Status and Student Group

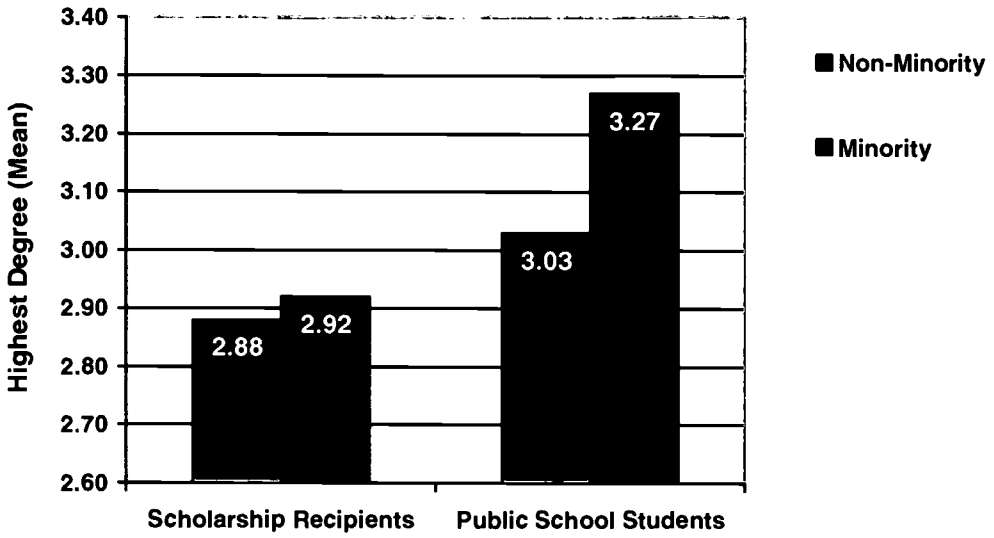


Teacher's Highest Degree Completed as function of Minority Status and Student Group

A 2 (minority status: minority vs. non-minority) by 2 (student group: scholarship recipients vs. public school students) between-subjects factorial analysis of variance was conducted using 2000-2001 "teacher's highest degree earned" (i.e., highest degree) as the dependent variable. Figure 4 displays the highest degree completed as a function both of minority status and of student group membership.

The analysis of variance revealed statistically significant main effects of minority status ($F[1, 1618] = 4.31$, $p = .038$) and student group ($F[1, 1618] = 13.71$, $p < .001$). Regardless of student group membership, minority students were taught by teachers with significantly more education and training ($M = 3.09$) than were non-minority students ($M = 2.95$). Furthermore, regardless of minority status, public school students were taught by teachers with significantly more education and training ($M = 3.15$) than were scholarship recipients in private schools ($M = 2.90$).

Figure 4. 2000-2001 Highest Degree Earned as a Function of Minority Status and Student Group



The student group by minority status interaction, however, was not statistically significant. The non-significant interaction indicates that, both in public and in private schools (among scholarship recipients), minority students experienced classes led by teachers with slightly more education and training than did non-minority students. In other words, the difference in the teachers' highest degree completed between minority and non-minority students did not vary as a function of student group membership (public school vs. private school scholarship recipients).

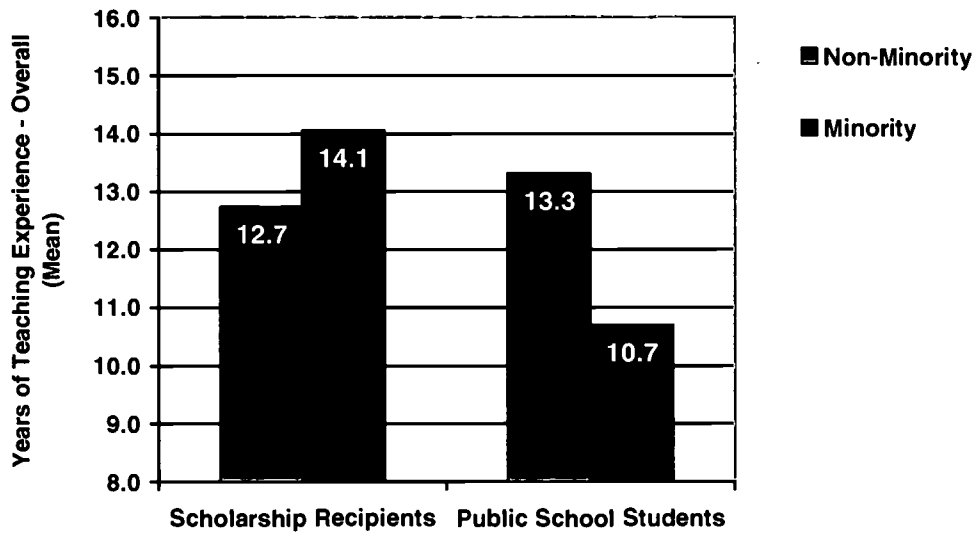
Years of Teaching Experience as function of Minority Status and Student Group

A 2 (minority status: minority vs. non-minority) by 2 (student group: scholarship recipients vs. public school students) between-subjects factorial analysis of variance was conducted using 2000-2001 years of teaching experience as the dependent variable. Figure 5 displays the mean years of teaching experience as a function of both minority status and student group membership.

The analysis of variance revealed a statistically significant main effect of student group ($F[1, 1641] = 5.34$, $p = .021$), but the main effect of minority status was not statistically significant. The non-significant main effect of minority status indicates that when student group membership is ignored, minority students' teachers did not possess significantly more years of experience than did non-minority students' teachers. In contrast, regardless of minority status, scholarship

recipients had teachers with significantly more experience (M = 13.4 years) than did public school students (M = 12.0 years).

Figure 5. 2000-2001 Mean Years of Overall Teaching Experience as a Function of Minority Status and Student Group



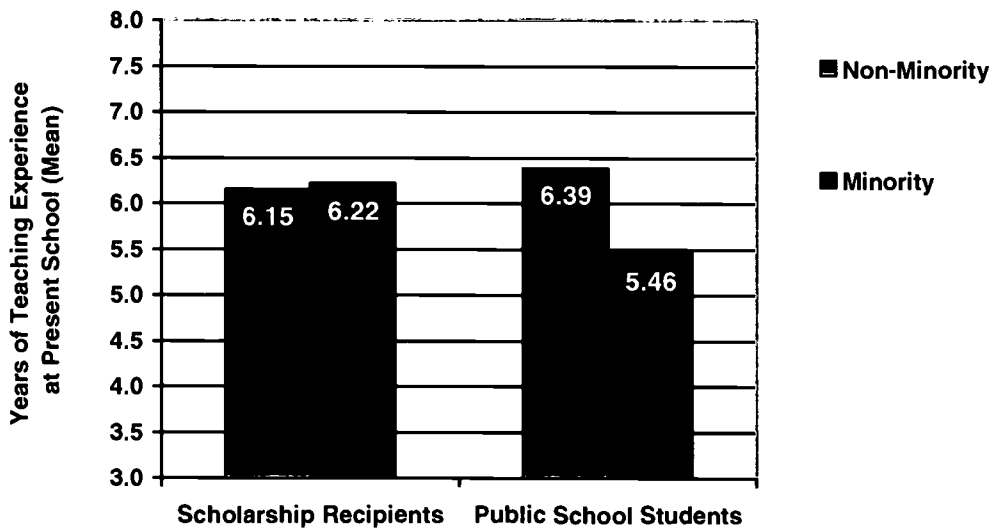
The main effect of student group membership was qualified by a statistically significant interaction between student group and minority status ($F[1, 1641] = 10.56$, $p = .001$), indicating that the difference in years of teaching experience between scholarship recipients and public school students varied depending on student minority status. Descriptively, minority scholarship students had teachers with more experience (M = 14.1 years) than did non-minority scholarship students (M = 12.7 years), whereas this pattern was reversed for public school students, such that minority public school students had teachers with less experience (M = 10.7 years) than did non-minority public school students (M = 13.3). Follow-up comparisons indicated that, within the scholarship recipient group, the difference in years of teaching experience between minority recipients and non-minority recipients was not statistically significant. Comparisons within the public school group, however, found that non-minority students attending public schools had teachers with significantly more years of experience than did minority students attending public schools.

Notably, minority scholarship recipients had teachers with significantly more years of experience (M = 14.1 years) than did minority students in public schools (M = 10.7 years). In contrast, the difference in years of teaching experience between non-minority scholarship recipients (M = 12.7) and non-minority public school students (M = 13.3) was not statistically significant. Thus, for non-minority students, it did not matter whether they were scholarship recipients in private schools or public school students because years of teaching experience did not differ between these groups. For minority students, however, years of teaching experience differed depending on their group status, such that minority scholarship recipients attended classes with teachers who possessed more experience than did the minority students' teachers in public schools (see Figure 5 above).

Years of Teaching Experience at Present School as function of Minority Status and Student Group

A 2 (minority status: minority vs. non-minority) by 2 (student group: scholarship recipients vs. public school non-applicant) between-subjects factorial analysis of variance was conducted using 2000-2001 years of teaching experience at present school as the dependent variable. Figure 6 displays the mean years of teaching experience at the present school as a function of both minority status and student group membership.

Figure 6. **2000-2001 Mean Years of Teaching Experience at Present School as a Function of Minority Status and Student Group**



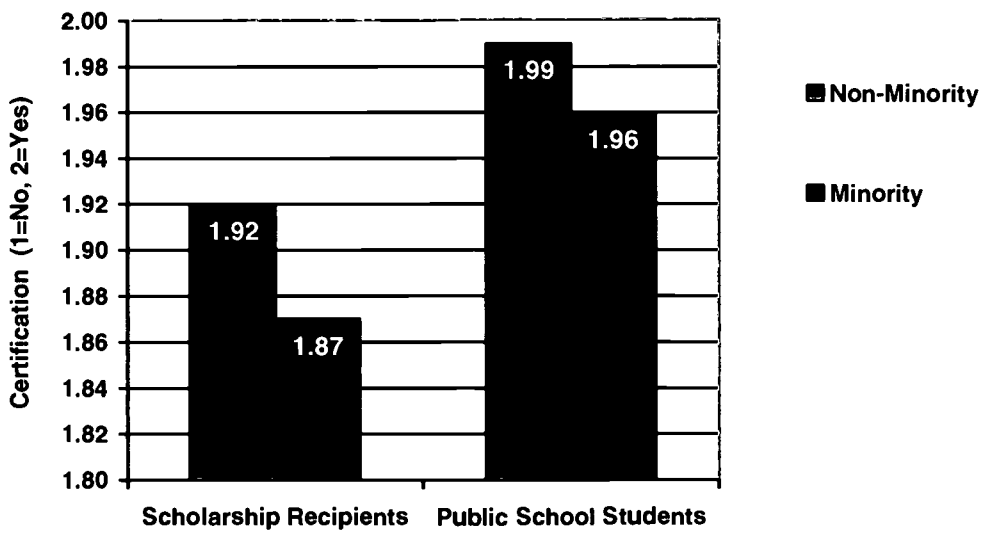
The analysis of variance, however, revealed neither the main effects nor the interaction were statistically significant. Therefore, the teacher's years at their present school did not differ as a function of minority status, student group, or the interaction of these two student-level variables.

Teacher's Certification as function of Minority Status and Student Group

A 2 (minority status: minority vs. non-minority) by 2 (student group: scholarship recipients vs. public school students) between-subjects factorial analysis of variance was conducted using 2000-2001 certification (1 = no, 2 = yes) as the dependent variable. Figure 7 displays the mean teaching certification as a function of both minority status and student group membership.

The analysis of variance revealed a statistically significant main effect of minority status ($F[1, 1628] = 8.29$, $p = .004$), a statistically significant main effect of student group ($F[1, 1628] = 29.43$, $p < .001$), but no statistically significant interaction between student group and minority status. Non-minority students had teachers who were more likely to be certified (95% certified) than were the teachers of minority students (91% certified), and this finding did not vary depending on whether students were scholarship recipients or public school students. Furthermore, scholarship recipients had teachers who were less likely to be certified (89% certified) than were the teachers of public school students (97% certified), and this finding did not vary as a function of minority status.

Figure 7. 2000-2001 Mean Teaching Certification as a Function of Minority Status and Student Group



Classroom Characteristics as a Function of Meal Code Status (Family Income)

Using estimated meal code status as a proxy for family income, a series of 2 (meal code status: free or reduced vs. paid) x 2 (student group: scholarship recipients vs. public school students) between-subjects factorial analysis of variance were conducted using 2000-2001 class size, teacher's highest degree earned, years of teaching experience, years of teaching experience at present school, and teacher's certification as the dependent variables. All 2000-2001 scholarship recipients were included in the analysis, regardless of when they entered the program. None of the analyses, however, found either a statistically significant main effect of meal code status or a statistically significant student group by meal code status interaction. That is, regardless of student group membership, the difference in each of the classroom/teacher variables between higher income students (meal code = paid) and lower income students (meal code = free or reduced) was not statistically significant.

3.2.4 Classroom Characteristics and Student Mobility

Using individual students as the unit of analysis, the number of schools changes that were made from kindergarten to third grade (i.e. student mobility) was regressed on the classroom and teacher characteristics. Specifically, the average class size, average years of teaching experience, average years of teaching experience at the teachers' present school, and the average highest degree earned by teachers in the classrooms to which students were exposed from kindergarten to third grade were simultaneously entered as predictors of mobility. As a set, the four-year average classroom and teacher variables accounted for a relatively small, but statistically significant, portion of the variance in student mobility ($F[4, 640] = 2.46$, $R^2 = .015$, $p = .045$). However, after controlling for the other classroom variables in the regression model (i.e., after statistically accounting for the common core of variance in mobility shared by the set of classroom variables), none of the individual classroom variables accounted for a statistically significant unique portion of the variance in student mobility.

Without any other classroom variables in the analysis, however, the students' average class size was found to have a weak, but statistically significant, inverse relationship with student mobility ($r = -.07$, $r^2 = .005$, $p = .013$). This indicates that students exposed to larger class sizes tended to make fewer changes in the schools they attended between kindergarten and third grade, whereas

students exposed to smaller classes tended to make more school changes. Similarly, both the average years of overall teaching experience ($r = -.09$, $r^2 = .008$, $p = .003$) and the average years of teaching experience at the teachers' present schools ($r = -.10$, $r^2 = .01$, $p = .004$) independently demonstrated statistically significant inverse relationships with student mobility. That is, students who were taught by teachers with more experience – both overall and at their current school at the time – tended to make fewer school changes between kindergarten and third grade, whereas students who were exposed to teachers with less experience tended to make more school changes during this time period.[42] The average highest degree earned by teachers in classrooms that students attended from kindergarten to third grade, however, was not independently related to student mobility at a statistically significant level ($r = -.009$, $r^2 < .0001$, $p = .391$).

[42] The simple correlation between class size and years of teaching experience indicates that teachers with more experience tend to teach larger classes ($r = .16$, $r^2 = .026$, $p = .004$). The nature of the present data, however, *does not* permit correlation to imply causation. Therefore, these analyses cannot determine the directionality of the associations among student mobility, teaching experience, and class size. For instance, class size may have a casual effect on student mobility, such that teaching experience is only indirectly associated with mobility via the relationship between teaching experience and class size. Alternatively, teaching experience may have a casual effect on mobility, such that class size is only indirectly associated with mobility. Which scenario is correct cannot be determined via this correlational analysis.

3.3 Question Three

What is the impact of participation in the Cleveland Scholarship and Tutoring Program on student academic achievement?

Comparative analyses of student achievement data collected across the four testing episodes (early and late first grade, late second grade, and late third grade) were conducted to investigate whether participation in the Cleveland Scholarship and Tutoring Program (CSTP) had an impact on academic achievement. To examine not only differences in achievement between and among the targeted groups of students, but also to investigate patterns in performance across time, mixed-design analyses of covariance were used in this portion of the project. Factorial mixed-design analyses included one between-subjects variable (student group) and one within-subject variable (testing episode). The results of these analyses allow examination of differences associated with: (a) time of testing, (b) group membership, and (c) the interaction of time of testing and group membership – to examine whether the pattern of change in academic achievement across time differed as a function of group membership.

Separate univariate analyses were conducted on each of the four academic measures provided by the Terra Nova across the four testing episodes (reading, language, mathematics, and total). Each analysis included three-year and four-year scholarship recipients, applicant non-recipients, non-applicants, and scholarship winner non-users.[43] Because student demographic characteristics accounted for a statistically significant, albeit small, portion of the variance in student achievement within each of the four testing episodes, student sex, minority status, and estimated meal code were included as covariates in the analyses. The classroom and teacher characteristics, however, failed to explain a substantial portion of variance in student achievement[44] or achievement change across the four testing episodes. Therefore, the present examination of longitudinal achievement data does not include classroom or teacher variables as covariates. In

[43] Analyses including achievement data collected from scholarship recipients who entered the program as second graders in 1999-2000 (two-year recipients) and as third graders in 2000-2001(one-year recipients) are included in separate analyses examining the effect of differential entry on academic achievement. The results of these analyses are presented in a subsequent section of the report.

[44] Student and classroom/teacher characteristic did not account for a statistically significant portion of the variance in 2000-2001 achievement scores or variance in achievement scores from any of the previous testing episodes (see Metcalf, K.K., *Cleveland Scholarship and Tutoring Program Evaluation: 1998-2000 Technical Report*, Bloomington, Indiana, Indiana Center for Evaluation, 2001).

addition, because missing data have been handled differently than in previous years of the project, the findings of the present analyses may be slightly discrepant from the findings reported in previous years.[45]

3.3.1 Reading Achievement

Descriptive statistics on students' adjusted reading achievement across the four testing episodes are presented in Table 23. Figure 8 presents these data graphically. The results of the mixed-design analysis of covariance are displayed in Table 24.[46]

A statistically significant main effect of time (i.e., testing episode) was found, and follow-up comparisons revealed that, regardless of student group membership, all students improved significantly across the testing episodes. That is, independent of group membership, student reading achievement was statistically higher at each successive testing period than it was previously (see Table 23 or Figure 8). The main effect of student group membership, however, was not statistically significant. In other words, after averaging across the four testing periods, reading achievement did not differ significantly among the student groups.

However, a statistically significant interaction was present between group membership and time of testing. To further examine how reading achievement across time varied as a function of group membership, follow-up pairwise comparisons were performed.[47] The results of these comparisons are depicted graphically in Figure 9.

[45] For the present report, missing data have been imputed (estimated and replaced), whereas prior evaluations eliminated cases with incomplete data and analyzed only complete cases (i.e., a listwise deletion technique was used is the past).

[46] The assumption of sphericity is likely to be violated with the univariate repeated measures or mixed-design analysis of variance. The sphericity assumption is violated when there is substantial dependence in observations across blocks within one or more participants that is due to random error (i.e., the residual error term for a participant is dependent across blocks). Although, Mauchly's Test of Sphericity ($W[5] = .90$, $p < .001$) revealed a violation of this assumption, estimates of sphericity for the analysis of reading scores still were strong (Greenhouse-Geisser = .999 - .942; Huynh-Feldt = .999 - .947). Therefore, no adjustments to degrees of freedom were made.

[47] The issues of homogeneity of variance and differential sample size pose problems for paired-comparison procedures. *Intransitive decisions* (e.g., outcomes suggesting mean A = mean B, mean B = mean C, but mean A < mean C) can occur using paired comparison procedures because they entail a series of discrete, pairwise significance tests (see Kirk, 1998 and others).

Table 23. Student Achievement in Reading: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

Student Group	N	Fall 1st Grade (1998) Mean	Fall 1st Grade (1998) SE	Spring 1st Grade (1999) Mean	Spring 1st Grade (1999) SE	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student Group) Mean	Marginal Means (Student Group) SE
3-year Scholarship Recipients	112	**546.21**	3.16	**564.86**	3.45	**595.36**	3.18	**609.17**	3.62	**578.90**	2.61
4-year Scholarship Recipients	339	**543.23**	1.84	**560.95**	2.01	**595.35**	1.85	**615.52**	2.11	**578.76**	1.52
Public Applicant Non-Recipients	376	**537.83**	1.72	**565.23**	1.88	**597.33**	1.74	**614.19**	1.97	**578.65**	1.42
Public Winner Non-Users	74	**538.64**	3.90	**570.51**	4.26	**597.56**	3.93	**615.61**	4.47	**580.58**	3.22
Public Non-Applicants	460	**531.97**	1.56	**563.14**	1.70	**595.26**	1.57	**613.49**	1.78	**575.97**	1.28
Marginal Means (Testing Episode)		**539.58**	1.16	**564.94**	1.27	**596.17**	1.17	**613.60**	1.33		

Means adjusted for the covariates included in the model: *sex* (M = 1.53), *minority status* (M = 1.76), and *estimated meal code* (M = 1.49).

Based on the test levels administered, the ranges of attainable reading scale scores are as follows: Fall 1st grade (355-626), Spring 1st grade (407-701), Spring 2nd grade (423-722), Spring 3rd grade (427-750)

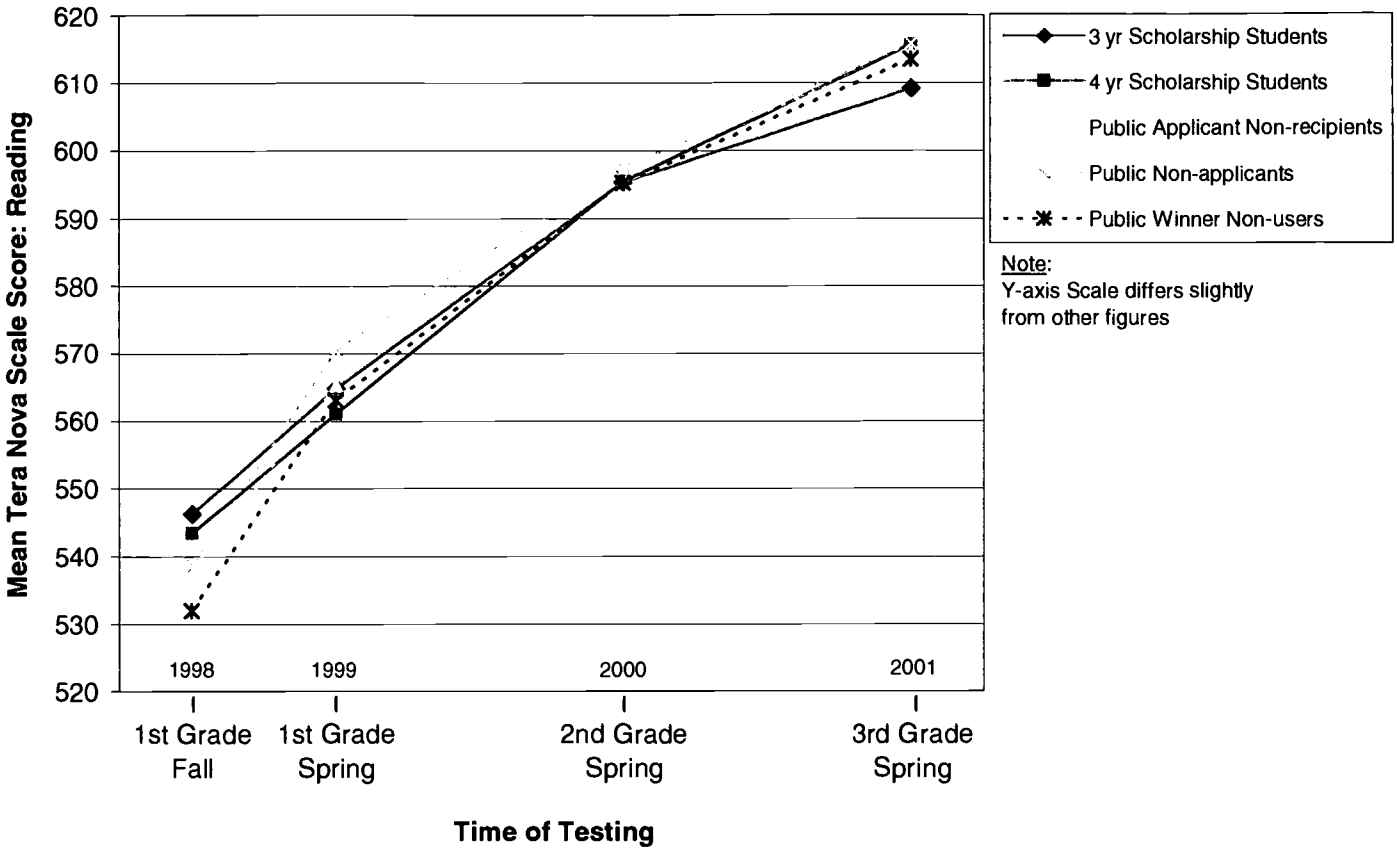
Table 24. Mixed Design Analysis of Covariance on Reading Achievement Scale Scores: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

Source of Variance[a]	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[b]
Within Subjects						
Time	502322.60	3	167440.87	252.453	< .001	.157
Time x Group	33052.28	12	2754.36	4.153	< .001	.012
Error	2692158.76	4059	663.25			
Between Subjects						
Group	11251.04	4	2812.76	.933	.444	< .001
Error	4079221.07	1353	3014.94			

a. Covariates not reported in table: sex (p<.001), time*sex (p = .022); minority status (p<.001), time*minority status (p =.037); meal code (p<.001), time*estimated meal code (p = .481 ns)

b. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 8. Reading Achievement: 1998 to 2001



At the beginning of first grade, both three-year (M = 546.21) and four-year scholarship recipients (M = 543.23) were achieving at significantly higher levels on the reading measure than were applicant non-recipients (M = 537.83) and non-applicants (M = 531.97). However, the difference in reading achievement between three-year and four-year scholarship recipients was not statistically significant, and neither scholarship recipient group differed significantly from winner non-users (M = 538.64) in reading achievement at the beginning of first grade. Among the public school comparison groups at the beginning of first grade, applicant non-recipients (M = 537.83) were achieving at a statistically higher level than were non-applicants (M = 531.97), but neither of these differed significantly from winner non-users (M = 538.64) in reading achievement.[48]

[48] The smaller sample size and relatively larger standard error associated with public school scholarship winner non-users resulted in intransitive findings (i.e., discrepancies between descriptive and inferential mean differences among the student groups) within the early first grade testing episode. Thus, interpreting findings that involve public school scholarship winner non-users should be done cautiously.

Figure 9. Pairwise Mean Differences in Reading Achievement among Student Groups by Testing Episode

Early First Grade (Fall 1998)

3-year Scholarship Recipients	4-year Scholarship Recipients	Winner Non-users	Applicant Non-Recipients	Non-Applicants
				531.97
			537.83	
		538.64		
546.21	543.23			

Late First Grade (Spring 1999)

Winner Non-users	Applicant Non-Recipients	3-year Scholarship Recipients	Non-Applicants	4-year Scholarship Recipients
	565.23	564.86	563.14	560.95
570.51				

Late Second Grade (Spring 2000)

Winner Non-users	Applicant Non-Recipients	3-year Scholarship Recipients	4-year Scholarship Recipients	Non-Applicants
597.56	597.33	595.36	595.35	595.26

Late Third Grade (Spring 2001)

Winner Non-users	4-year Scholarship Recipients	Applicant Non-Recipients	Non-Applicants	3-year Scholarship Recipients
615.61	615.52	614.19	613.49	609.17

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

By the end of first grade, no statistically significant differences in reading achievement were found among three-year scholarship recipients (M = 564.86), four-year scholarship recipients (M = 560.95), applicant non-recipients (M = 565.23), and non-applicants (M = 563.14). By the end of first grade, these groups of students were achieving at statically similar levels on the reading measure. Notably, winner non-users (M = 570.51) were achieving at a statistically higher level than were four-year scholarship recipients at the end of first grade (M = 560.95). Descriptively, by the end of first grade, winner non-users were achieving at the highest level in reading, whereas four-year scholarship recipients were achieving at the lowest level—in comparison to the other student groups in the analysis. By the end of second grade and again by the end of third grade, however, no statistically significant differences in reading achievement were found among the student groups.

To further explicate this interaction, follow-up analyses examined the change in scores for each group from one testing episode to another. Specifically, a set of pairwise comparisons examined whether the amount of change in reading achievement that occurred within each group across time differed among the student groups. In other words, did one group's performance change more or less than did another group's performance from one testing episode to another, and between which testing episodes did the differential change in achievement occur? Because the main effect of time indicates that reading achievement increased across time, regardless of group membership, the following analyses examine whether one group of students *gained* more or less in achievement (i.e., experienced greater or less improvement) across the testing episodes than did another group of students. Furthermore, the interaction comparisons not only analyze differential change in performance between the student groups sequentially from time one (early first grade) to time two (late first grade), from time two to time three (late second grade), and from time three to time four (late third grade), but the comparisons also analyze overall change from time one (early first grade) to time four (late third grade) as well as from time two (late first grade) to time four (late third grade).

The results of the interaction comparisons of achievement score change are graphically presented in Figure 10. From early first grade to late first grade, the interaction comparisons found that all of the public school comparison groups gained more in reading achievement than did the four-

year scholarship recipients. Specifically, the group of students who entered the CSTP as kindergartners in 1997, and have remained in the program for four years, demonstrated a statistically *smaller* average increase in reading achievement scores (M = 17.72)[49] from the beginning of first grade to the end of first grade than did: applicant non-recipients (M = 27.4); non-applicants (M = 31.17); or winner non-users (M = 31.87). In addition, non-applicants displayed a greater increase in reading achievement from early first grade to late first grade than did the three-year scholarship recipients (M = 18.65).

From late first grade (1999) to late second grade (2000), as well as from late second grade (2000) to late third grade (2001), no statistically significant differences were found among the comparison groups in terms of the amount of reading achievement gain that occurred within each group of students across time. The five groups demonstrated statistically similar improvement in reading achievement from late first grade to late second grade (three year scholarship recipients [M = 30.50], four-year scholarship recipients [M = 34.40], applicant non-recipients [M = 32.10], non-applicants [M = 32.12], and winner non-users [M = 27.05]), and from late second grade to late third grade (three year scholarship recipients [M = 13.81], four-year scholarship recipients [M = 20.17], applicant non-recipients [M = 16.86], non-applicants [M = 18.23], and non-users [M = 18.05]).

Reading Achievement Summary. The statistically significant interaction, which indicates that the pattern of reading achievement across time differed as a function of student group, can be explained by: (a) statistically significant differences in initial reading achievement in early first grade, such that three- and four-year scholarship students initially outperformed all of the public school comparison groups, followed by (b) differential gain in reading achievement between the public school comparison groups and the scholarship recipients, such that by the end of first grade, all of the student groups were achieving at a statically similar level on the reading test (with the exception of public school scholarship winner non-users who had caught up to and surpassed four-year scholarship recipients). After the late first grade testing episode, the student groups not only displayed statistically similar patterns of change in reading achievement across time, but the groups also performed at statistically similar levels both within the second grade testing episode and within the third grade testing episode.

[49] Throughout this report, the symbol, M (M-delta) is used to denote "mean change" in achievement scores.

Figure 10. Pairwise Mean Differences in Reading Achievement CHANGE among Student Groups by Testing Episode

Early First Grade (Fall 1998) to Late First Grade (Spring 1999): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	3-year Scholarship Recipients	4-year Scholarship Recipients
			18.65	17.72
	31.17	27.40		
31.87				

Late First Grade (Spring 1999) to Late Second Grade (Spring 2000): *Scale Score Change*

4-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	3-year Scholarship Recipients	Winner Non-users
34.40	32.12	32.10	30.50	27.05

Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001): *Scale Score Change*

4-year Scholarship Recipients	Non-Applicants	Winner Non-users	Applicant Non-Recipients	3-year Scholarship Recipients
20.17	18.23	18.05	16.86	13.81

Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): *Scale Score Change*

Non-Applicants	Winner Non-users	Applicant Non-Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients
			72.29	62.96
	76.97	76.36		
81.52				

Late First Grade (Spring 1999) to Late Third Grade (Spring 2001): *Scale Score Change*

4-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	Winner Non-users	3-year Scholarship Recipients
54.57	50.35	48.96	45.10	44.31

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

3.3.2 Language Achievement

Descriptive statistics on students' adjusted language achievement across the four testing episodes are presented in Table 25. Figure 11 presents these data graphically. The results of the mixed-design analysis of covariance are displayed in Table 26.[50]

As in reading achievement, a statistically significant main effect of time (testing episode) was found, and follow-up comparisons revealed that, regardless of student group membership, all students improved significantly across the testing episodes. The main effect of student group membership, however, was not statistically significant. In other words, after averaging across the four testing periods, language achievement did not differ significantly among the student groups.

More important to the present discussion, a statistically significant interaction was present between group membership and time of testing indicating that the pattern of change in language achievement across time differed between at least two of the student groups. Follow-up comparisons first were made among the groups within each testing episode. The results of these comparisons are depicted graphically in Figure 12.

Table 25. Student Achievement in Language: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

		Fall 1st Grade (1998)		Spring 1st Grade (1999)		Spring 2nd Grade (2000)		Spring 3rd Grade (2001)		Marginal Means (Student Group)	
Student Group	N	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
3-year Scholarship Recipients	112	**549.81**	3.29	**565.87**	3.51	**589.41**	3.39	**610.34**	3.13	**578.86**	2.64
4-year Scholarship Recipients	339	**548.26**	1.91	**562.22**	2.04	**590.87**	1.98	**612.37**	1.82	**578.43**	1.54
Public Applicant Non-Recipients	376	**539.93**	1.79	**563.11**	1.92	**584.69**	1.85	**608.46**	1.71	**574.05**	1.44
Public Winner Non-Users	74	**533.55**	4.06	**564.14**	4.33	**591.57**	4.19	**606.72**	3.87	**573.99**	3.26
Public Non-Applicants	460	**538.77**	1.62	**563.06**	1.73	**585.51**	1.67	**609.08**	1.54	**574.10**	1.30
Marginal Means (Testing Episode)		**542.06**	1.20	**563.68**	1.28	**588.41**	1.24	**609.39**	1.14		

Means adjusted for the covariates included in the model: sex (M = 1.53), minority status (M = 1.76), and estimated meal code (1.49)

Based on the test levels administered, the ranges of attainable language scale scores are as follows: Fall 1st grade (325-620), Spring 1st grade (400-680), Spring 2nd grade (424-706), Spring 3rd grade (455-730).

[50] Although, Mauchly's Test of Sphericity (W[5] = .957, $p < .001$) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of language scores still were strong (Greenhouse-Geisser = .999 - .973; Huynh-Feldt = .999 - .980). Therefore, no adjustments to degrees of freedom were made.

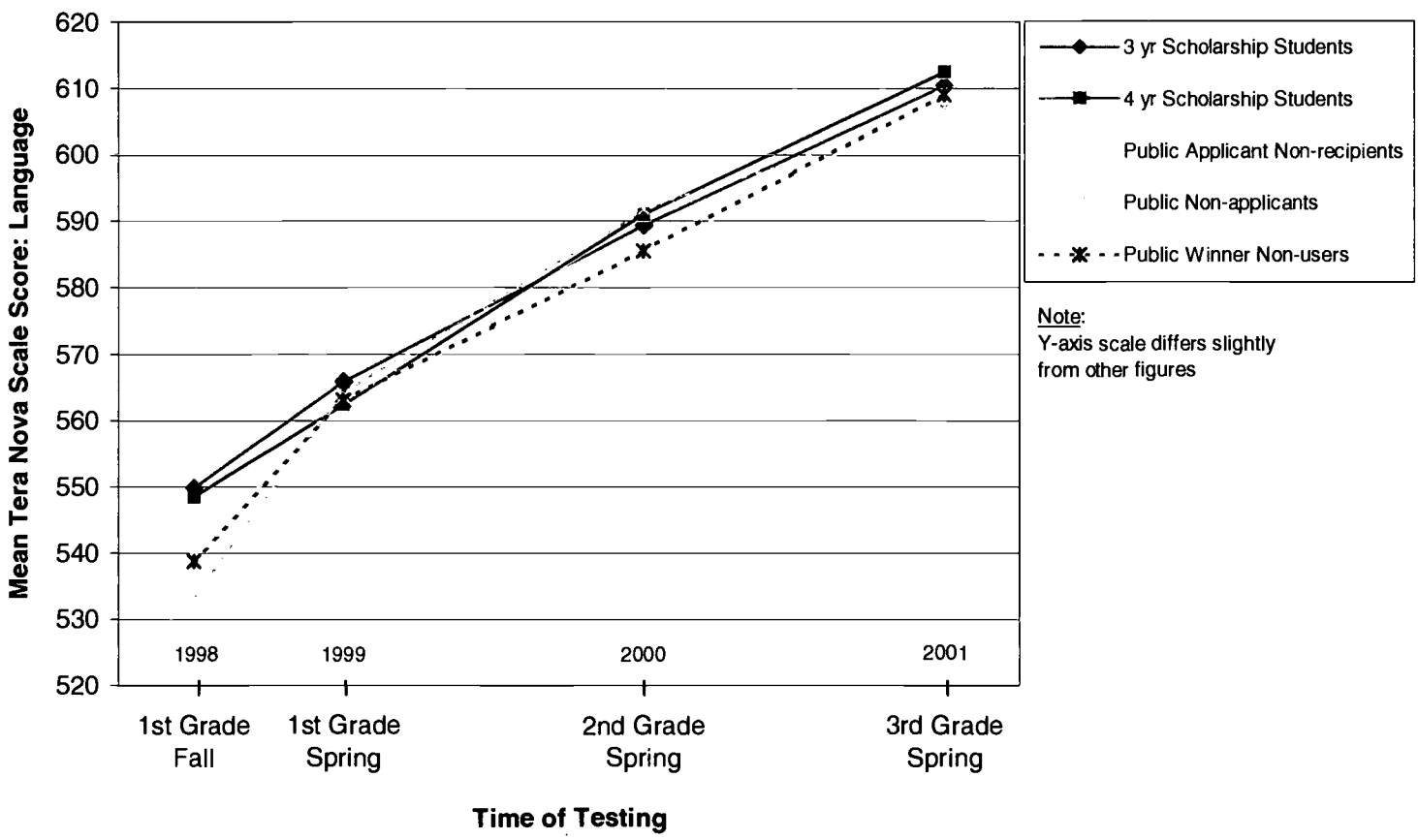
Table 26. Mixed Design Analysis of Covariance on Language Achievement Scale Scores: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

Source of Variance[a]	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[b]
Within Subjects						
Time	76756.51	3	25585.50	41.608	< .001	.030
Time x Group	21868.19	12	1822.35	2.964	< .001	< .001
Error	2495972.80	4059	614.92			
Between Subjects						
Group	22711.72	4	5677.93	1.838	.119	< .001
Error	4178569.47	1353	3088.37			

a. Covariates not reported in table: sex (p<.001), time*sex (p=.122 ns); minority status (p<.001), time*minority status (p=.040); meal code (p<.001), time*meal code (p=.525 ns)

b. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 11. Language Achievement: 1998 to 2001



These analyses revealed that both three-year and four-year scholarship recipients were achieving at statistically higher levels on the language test than were students in any of the public school comparison groups at the beginning of first grade. Specifically, in the early first grade testing episode, mean language achievement scores for the following comparison groups were significantly lower than the mean language achievement scores for three-year (M = 549.81) and four-year scholarship recipients (M = 548.26): public school applicant-non-recipients (M = 539.93), public school non-applicants (M = 538.77), and public school scholarship winner non-users (M = 533.55). No statistically significant differences were found between three-year and four-year scholarship recipients or among the three public school groups.

By the end of first grade, there were no significant differences in language achievement among the groups. By the end of the second grade, however, statistically significant differences in language achievement again were found among the comparison groups. Specifically, by the end of second grade, four-year scholarship recipients (M = 590.87) were achieving at a statistically higher level on the language measure than were either applicant non-recipients (M = 584.69) or non-applicants (M = 563.06). No statistically significant differences in language achievement were found between three-year scholarship recipients (M = 589.41) and the other student groups at the end of second grade. Moreover, no statistically significant differences in language achievement were found among the three public school comparison groups (see Figure 12).

By the end of third grade, no statistically significant differences in language achievement were found among the groups of students. That is, three-year and four-year scholarship recipients did not differ significantly in language achievement at the end of third grade; neither three-year nor four-year scholarship recipients differed significantly from any of the public school comparison groups; and no statistically significant differences in language achievement were revealed among the three public school comparison groups at the end of third grade (see Figure 12).

To further examine the nature of the student group by testing episode interaction, follow-up comparisons examined differences in the amount of change in language achievement that occurred within each group from one testing episode to another. The results of these comparisons are depicted graphically in Figure 13.

Figure 12. Pairwise Mean Differences in Language Achievement among Student Groups by Testing Episode

Early First Grade (Fall 1998)

3-year Scholarship Recipients	4-year Scholarship Recipients	Applicant Non-Recipients	Non-Applicants	Winner Non-users
		539.93	538.77	533.55
549.81	548.26			

Late First Grade (Spring 1999)

3-year Scholarship Recipients	Winner Non-users	Applicant Non-Recipients	Non-Applicants	4-year Scholarship Recipients
565.87	564.14	563.11	563.06	562.22

Late Second Grade (Spring 2000)

Winner Non-users	4-year Scholarship Recipients	3-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients
			585.51	584.69
	590.87			
591.57		589.41		

Late Third Grade (Spring 2001)

4-year Scholarship Recipients	3-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	Winner Non-users
612.37	610.34	609.08	608.46	606.72

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

Figure 13. Pairwise Mean Differences in Language Achievement CHANGE among Student Groups by Testing Episode

Early First Grade (Fall 1998) to Late First Grade (Spring 1999): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	3-year Scholarship Recipients	4-year Scholarship Recipients
			16.06	13.96
	24.29	23.18		
30.59				

Late First Grade (Spring 1999) to Late Second Grade (Spring 2000): *Scale Score Change*

4-year Scholarship Recipients	Winner Non-users	3-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients
			22.45	21.58
	27.43	23.54		
28.65				

Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001): *Scale Score Change*

Applicant Non-Recipients	Non-Applicants	4-year Scholarship Recipients	3-year Scholarship Recipients	Winner Non-users
23.77	23.57	21.50	20.93	15.15

Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients
		68.53	64.11	60.53
	70.31			
73.17				

Late First Grade (Spring 1999) to Late Third Grade (Spring 2001): *Scale Score Change*

4-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	3-year Scholarship Recipients	Winner Non-users
50.15	46.02	45.35	44.47	42.58

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

From early first grade to late first grade, the interaction comparisons found that the increase in language achievement for each of the public school comparison groups was statistically greater than the gain experienced by four-year scholarship recipients. Applicant non-recipients (M = 23.18), non-applicants (M = 24.29), and winner non-users (M = 30.59) each demonstrated a greater gain in language achievement from early to late first grade than did the four-year scholarship recipients (M = 13.96). Furthermore, from early to late first grade, winner non-users (M = 30.59) also gained significantly more in language achievement than did three-year scholarship recipients (M = 16.06). No other statistically significant differences in language achievement gain from the beginning to the end of first grade were found among the student groups.

From the end of first grade to the end of second grade, four-year scholarship recipients (M = 28.65) demonstrated a statistically greater gain in language achievement than did applicant non-recipients (M = 21.58) and non-applicants (M = 22.45). However, the difference in language achievement gain between four-year scholarship recipients and both three-year scholarship recipients (M = 23.54) and winner-non users (M = 27.43) was not statistically significant. Moreover, no statistically different gains in language achievement were revealed either among the public school comparison groups or between the three-year scholarship recipients and any of the public school comparison groups during second grade.

Finally, from the end of second grade to the end of third grade, all of the student groups displayed statistically similar gains in language achievement: three year scholarship recipients (M = 20.93), four-year scholarship recipients (M = 21.50), applicant non-recipients (M = 23.77), non-applicants (M = 23.57), and winner non-users (M = 15.15).

Total gain in language achievement from early first grade to late third grade was statistically greater for public school non-applicants (M = 70.31) than it was for either three-year scholarship recipients (M = 60.53) or four-year scholarship recipients (M = 64.11).[51] No other statistically significant differences in language achievement gain were found among the groups from the beginning of first grade to the end of third grade (public school applicant non-

[51] Analyses of achievement gain from early first grade (autumn, 1998) and late first grade (spring, 1999) to late third grade (spring, 2001) take into account only the level of achievement in those two testing episodes, ignoring any differential gain that occurred between interim testing episodes.

recipients, M = 68.53; public school scholarship winner non-users, M = 73.17). In addition, from the end of first grade to the end of third, all of the groups demonstrated statistically similar gains in language achievement: three year scholarship recipients (M = 44.47), four-year scholarship recipients (M = 50.15), applicant non-recipients (M = 45.35), non-applicants (M = 46.02), and winner non-users (M = 42.58).

Language Achievement Summary. After entering first grade with significantly lower language achievement than three-year and four-year scholarship recipients, public school students demonstrated greater gain in language achievement during first grade than either of the two scholarship groups. Consequently, all of the groups were achieving at statistically similar levels by the end of first grade. From late first grade to late second grade, four-year (but not three-year) scholarship recipients demonstrated greater average gains in language achievement than did the public school comparison students, with the exception of winner non-users. Therefore, by the end of second grade, four-year scholarship recipients again were achieving at statistically higher levels on the language measure than were the public school comparison groups. Finally, from the end of second grade to the end of third grade, all of the groups improved a statistically similar amount, with public school students improving sufficiently to eliminate any significant differences between public school and scholarship students by spring of their third grade year.

3.3.3 Mathematics Achievement

Descriptive statistics on students' adjusted mathematics achievement across the four testing episodes are presented in Table 27. Figure 14 presents these data graphically. The results of the mixed-design analysis of covariance are displayed in Table 28.[52]

As with the reading and language achievement scores, a statistically significant main effect of time (testing episode) was found. Paired comparisons revealed that, regardless of group membership, all students improved significantly in mathematics across the four testing episodes. No significant main effect for group was indicated. In other words, after averaging across the four testing episodes, mathematics achievement did not differ significantly among the student

[52] Although, Mauchly's Test of Sphericity ($W[5] = .926$, $p < .001$) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of reading scores still were strong (Greenhouse-Geisser = .999 - .948; Huynh-Feldt = .999 - .956). Therefore, no adjustments to degrees of freedom were made.

groups. However, a statistically significant interaction was present between group membership and time of testing.

Table 27. Student Achievement in Mathematics: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

Student Group	N	Fall 1st Grade (1998) Mean	Fall 1st Grade (1998) SE	Spring 1st Grade (1999) Mean	Spring 1st Grade (1999) SE	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student Group) Mean	Marginal Means (Student Group) SE
3-year Scholarship Recipients	112	**499.84**	3.01	**517.42**	3.13	**548.84**	3.24	**582.95**	3.52	**537.26**	2.67
4-year Scholarship Recipients	339	**499.35**	1.75	**517.21**	1.82	**548.76**	1.88	**591.52**	2.05	**539.21**	1.55
Public Applicant Non-Recipients	376	**490.10**	1.64	**515.17**	1.71	**550.71**	1.77	**592.12**	1.92	**537.02**	1.46
Public Winner Non-Users	74	**489.78**	3.71	**518.71**	3.87	**552.64**	4.00	**594.73**	4.35	**538.97**	3.30
Public Non-Applicants	460	**491.32**	1.48	**517.84**	1.54	**553.72**	1.60	**594.03**	1.74	**539.23**	1.32
Marginal Means (Testing Episode)		**494.08**	1.10	**517.27**	1.14	**550.93**	1.18	**591.07**	1.29		

Means adjusted for the covariates included in the model: sex (M = 1.53), minority status (M = 1.76), and estimated meal code (M = 1.49)
Based on the test levels administered, the ranges of attainable mathematics scale scores are as follows: Fall 1st grade (290-629), Spring 1st grade (324-680), Spring 2nd grade (347-720), Spring 3rd grade (385-740)

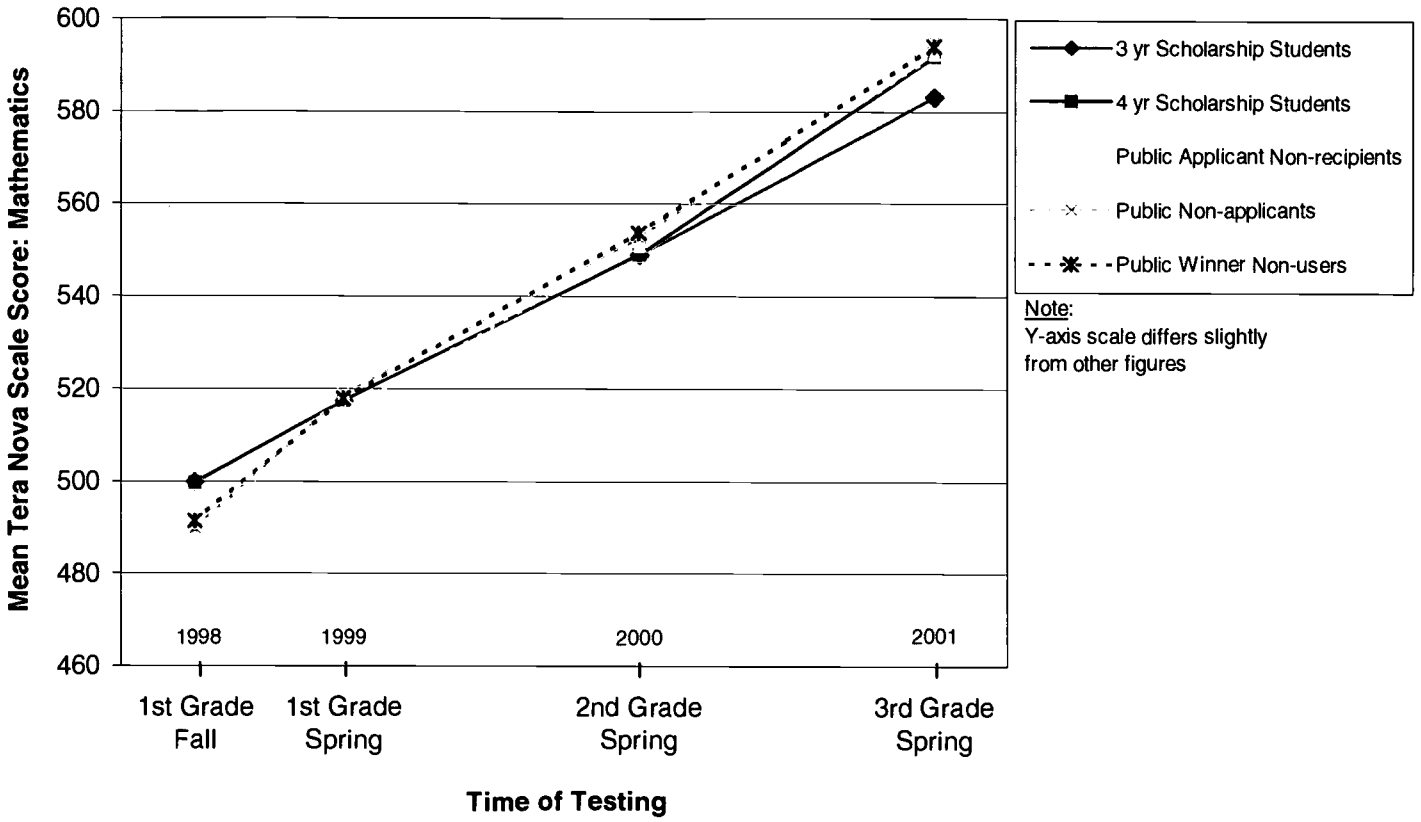
Table 28. Mixed Design Analysis of Covariance on Mathematics Achievement Scale Scores: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

Source of Variance[a]	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[b]
Within Subjects						
Time	186282.29	3	62094.10	126.648	< .001	.080
Time x Group	35444.98	12	2953.75	6.024	< .001	.010
Error	1990087.919	4059	490.29			
Between Subjects						
Group	5629.08	4	1407.27	.445	.776	< .001
Error	4279737.32	1353	3163.14			

a. Covariates not reported in table: sex (p=.076 ns), time*sex (p<.001); minority status (p<.001), time*minority status (p=.259 ns); meal code (p<.001), time*meal code (p=.267 ns)

b. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 14. Mathematics Achievement: 1998 to 2001



To further examine how mathematics achievement varied across time as a function of group membership, follow-up paired comparisons were performed. Comparisons first were made among the groups within each testing episode. The results of these comparisons are depicted graphically in Figure 15.

Comparisons revealed that mean mathematics achievement for three-year (M = 499.84) and four-year scholarship recipients (M = 499.35) was significantly higher than for applicant-non-recipients (M = 490.10), non-applicants (M = 491.32); and winner non-users (M = 489.78). No statistically significant differences were found among the scholarship recipients or among the three public school comparison groups.

Figure 15. Pairwise Mean Differences in Mathematics Achievement among Student Groups by Testing Episode

Early First Grade (Fall 1998)

3-year Scholarship Recipients	4-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	Winner Non-users
		491.32	490.10	489.78
499.84	499.35			

Late First Grade (Spring 1999)

Winner Non-users	Non-Applicants	3-year Scholarship Recipients	4-year Scholarship Recipients	Applicant Non-Recipients
518.71	517.84	517.42	517.21	515.17

Late Second Grade (Spring 2000)

Non-Applicants	Winner Non-users	Applicant Non-Recipients	3-year Scholarship Recipients	4-year Scholarship Recipients
	552.64	550.71	548.84	548.76
553.72				

Late Third Grade (Spring 2001)

Winner Non-users	Non-Applicants	Applicant Non-Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients
				582.95
594.73	594.03	592.12	591.52	

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

By the end of the first grade, no statistically significant differences in mathematics achievement were found among the student groups but, by the end of second grade, significant differences had emerged among the groups. Specifically, by the end of second grade, non-applicants (M = 553.72) were achieving at a statistically higher level in mathematics than were four-year

scholarship recipients (M = 548.76). No other statistically significant differences in mathematics achievement were found among the student groups at the end of second grade (see Figure 15).

By the end of third grade, however, four-year scholarship recipients (M = 591.52), applicant non-recipients (M = 592.12), non-applicants (M = 594.03), and winner non-users (M = 594.73) all were achieving at significantly higher levels in mathematics than were three-year scholarship recipients (M = 582.95). No additional statistically significant differences in mathematics achievement were identified.

To further examine the nature of the group by testing episode interaction, paired comparisons examined differences in the amount of change in mathematics achievement that occurred within each student group from one testing episode to another. The results of these comparisons are depicted graphically in Figure 16. From early first grade to late first grade, applicant non-recipients (M = 25.07), non-applicants (M = 26.52), and winner non-users (M = 28.93) each demonstrated significantly greater gain in mathematics achievement than did four-year scholarship recipients (M = 17.86). In addition, non-applicants (M = 26.52) displayed a statistically greater gain in mathematics during this period than did three-year scholarship recipients (M = 17.58). No other statistically significant differences in mathematics achievement gain from early to late first grade were found among the groups.

From the end of first grade to the end of second grade, no statistically significant differences in mathematics achievement gain were found among the student groups. That is, all of the groups displayed a statistically similar gain in mathematics achievement during this year: three-year scholarship recipients (M = 31.42), four-year scholarship recipients (M = 31.55), applicant non-recipients (M = 35.54), non-applicants (M = 35.88), and winner non-users (M = 33.93). A pattern of statistically similar gains in mathematics achievement continued from the end of second grade to the end of third grade: Three-year scholarship recipients (M = 34.11), four-year scholarship recipients (M = 42.76), public school applicant non-recipients (M = 41.41), public school non-applicants (M = 40.31), and public school scholarship winner non-users (M = 42.09).

Figure 16. Pairwise Mean Differences in Mathematics Achievement CHANGE among Student Groups by Testing Episode

Early First Grade (Fall 1998) to Late First Grade (Spring 1999): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients
			17.86	17.58
		25.07		
28.93	26.52			

Late First Grade (Spring 1999) to Late Second Grade (Spring 2000): *Scale Score Change*

Non-Applicants	Applicant Non-Recipients	Winner Non-users	4-year Scholarship Recipients	3-year Scholarship Recipients
35.88	35.54	33.93	31.55	31.42

Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001): *Scale Score Change*

4-year Scholarship Recipients	Winner Non-users	Applicant Non-Recipients	Non-Applicants	3-year Scholarship Recipients
42.76	42.09	41.41	40.31	34.11

Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients
				83.11
			92.17	
104.95	102.71	102.02		

Late First Grade (Spring 1999) to Late Third Grade (Spring 2001): *Scale Score Change*

Applicant Non-Recipients	Non-Applicants	Winner Non-users	4-year Scholarship Recipients	3-year Scholarship Recipients
				65.53
			74.31	
		76.02		
76.95	76.19			

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

Across the period from early first grade to late third grade, total gain in mathematics achievement was statistically smaller for three-year scholarship recipients (M = 83.11) than it was for any of the other groups (four-year scholarship recipients [M = 92.17], applicant non-recipients [M = 102.02], non-applicants [M = 102.71], and winner non-users [M = 104.95]). Moreover, although four-year scholarship recipients gained statistically more in mathematics than did three-year scholarship recipients, their gains were significantly smaller than any of the three public school comparison groups. Total gain for these public school groups was not statistically different during first grade.

From the end of first grade to the end of third grade, four-year scholarship recipients (M = 74.31), applicant non-recipients (M = 76.95), and non-applicants (M = 76.19), each displayed greater gain in mathematics achievement than did the three-year scholarship recipients (M = 65.53). No additional significant differences in mathematics achievement gain from late first grade to late third grade were found among the student groups (winner non-users, M = 76.02).

Mathematics Achievement Summary. After initial differences in mathematics achievement at the beginning of first grade favoring scholarship recipients, public school students demonstrated greater gains in mathematics achievement during first grade than did the scholarship recipients. Consequently, all of the student groups were achieving at statistically similar levels by the end of first grade. From the end of first grade to the end of second grade, all of the student groups displayed statistically similar patterns of gain in mathematics achievement, and only the highest achieving group (public non-applicants) and the lowest achieving group (four-year scholarship recipients) differed at a statistically significant level by the end of the second grade. Finally, from the end of second grade to the end of third grade, three-year scholarship recipients gained less in mathematics achievement than did students in the other groups. Despite the fact that this differential gain was not statistically significant, all of the student groups were achieving at a statistically higher level in mathematics in comparison to three-year scholarship recipients at the end of third grade.

3.3.4 Total Achievement (Average of Reading, Language, and Mathematics Scale Scores)

Descriptive statistics on students' adjusted total achievement across the four testing episodes are presented in Table 29. Figure 17 presents these data graphically. The results of mixed-design analysis of covariance are displayed in Table 30.[53]

Table 29. Student Total Achievement: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

Student Group	N	Fall 1st Grade (1998)		Spring 1st Grade (1999)		Spring 2nd Grade (2000)		Spring 3rd Grade (2001)		Marginal Means (Student Group)	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
3-year Scholarship Recipients	112	**531.94**	2.59	**549.71**	2.81	**577.96**	2.83	**600.82**	3.04	**565.11**	2.44
4-year Scholarship Recipients	339	**530.38**	1.51	**546.73**	1.64	**578.45**	1.65	**606.46**	1.77	**565.51**	1.42
Public Applicant Non-Recipients	376	**522.61**	1.42	**547.83**	1.54	**577.37**	1.55	**604.93**	1.66	**563.19**	1.33
Public Winner Non-Users	74	**520.67**	3.20	**551.11**	3.47	**580.58**	3.50	**605.69**	3.76	**564.51**	3.01
Public Non-Applicants	460	**520.67**	1.28	**548.02**	1.39	**578.30**	1.40	**605.51**	1.50	**563.12**	1.20
Marginal Means (Testing Episode)		**525.25**	0.95	**548.68**	1.03	**578.53**	1.03	**604.68**	1.11		

Means adjusted for the covariates included in the model: sex (M = 1.53), minority status (M = 1.76), and estimated meal code (M = 1.49)
Based on the test levels administered, the ranges of attainable total scale scores are as follows: Fall 1st grade (323-625), Spring 1st grade (377-687), Spring 2nd grade (398-716), Spring 3rd grade (422-740)

Table 30. Mixed Design Analysis of Covariance on Total Achievement Scale Scores: Early First Grade (Fall 1998) to Late Third Grade (Spring 2001)

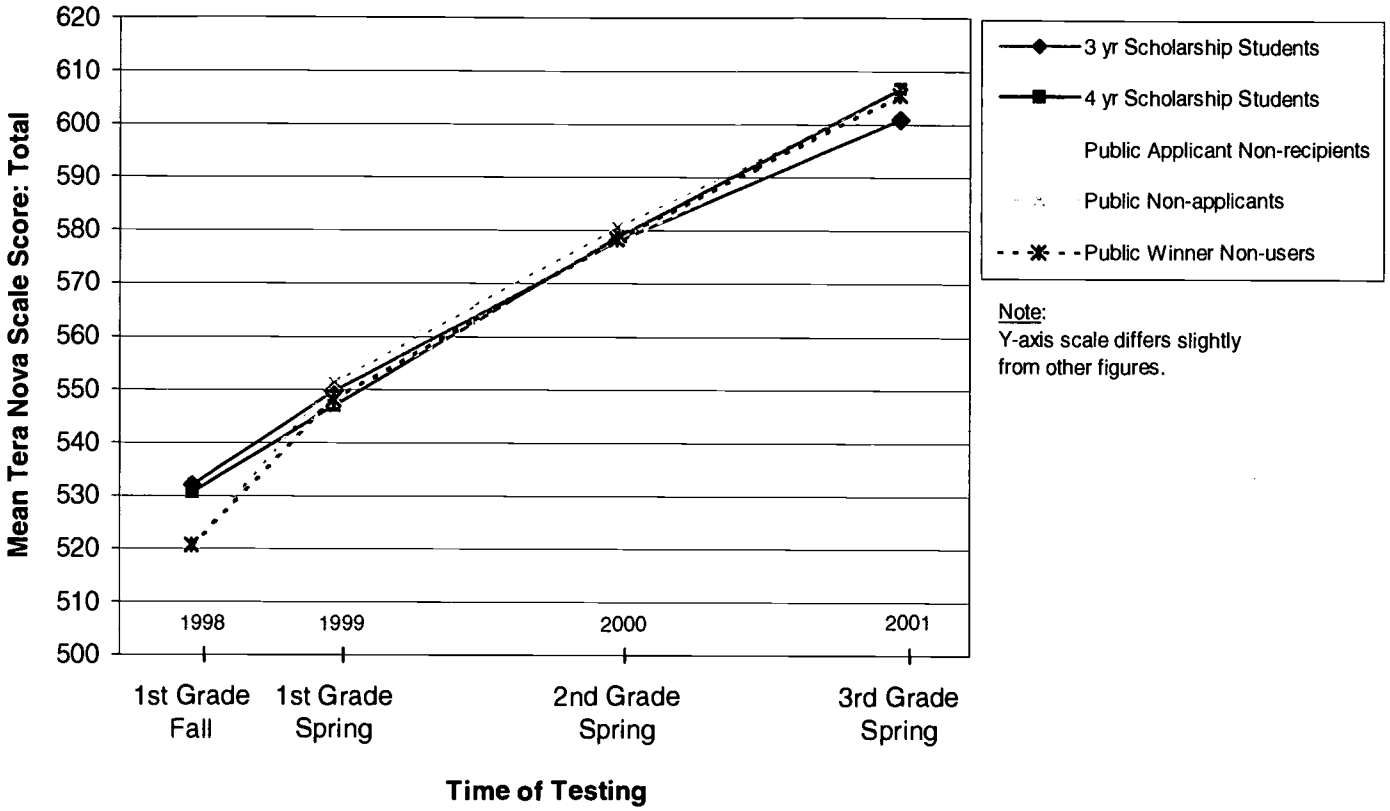
Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[b]
Within Subjects						
Time	107375.46	3	35791.82	118.137	< .001	.080
Time x Group	24802.83	12	2066.90	6.822	< .001	.020
Error	1229751.58	4059	302.96			
Between Subjects						
Group	5838.50	4	1459.63	.554	.696	< .001
Error	3562838.59	1353	2633.28			

a. Covariates not reported in table: sex (p<.001), time*sex (p<.001); minority status (p<.001), time*minority status (p=.025); meal code (p<.001), time*meal code (p=.113 ns)

b. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

[53] Although, Mauchly's Test of Sphericity (W[5] = .926, p < .001) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of reading scores still were strong (Greenhouse-Geisser = .999 - .948; Huynh-Feldt = .999 - .956). Therefore, no adjustments to degrees of freedom were made.

Figure 17. Total Achievement (Average of Reading, Language, and Mathematics): 1998 to 2001



As statistically significant main effect of time (testing episode) was found. Paired comparisons revealed that, regardless of student group membership, all students improved significantly in across the four testing episodes. That is, independent of group membership, total achievement was statistically higher at each successive testing episode than it was previously. No significant main effect for group was found; however, a significant interaction effect of testing episode and student group was indicated.

To further examine how total achievement across time varied as a function of group membership, follow-up paired comparisons were performed. The results of these comparisons are depicted graphically in Figure 18. Comparisons revealed that both three-year and four-year scholarship recipients were achieving at a statistically higher level than were students in any of the public school comparison groups at the beginning of first grade. Specifically, in the early first grade testing episode, the mean total achievement score for the following comparison groups was significantly lower than the mean total achievement score for three-year (M = 531.94) and four-year scholarship recipients (M = 530.38): applicant-non-recipients (M = 522.61), non-applicants

(M = 520.67); and winner non-users (M = 520.67). Follow-up comparisons, however, did not find a statistically significant difference in mean total achievement between three-year and four-year scholarship recipients at the beginning of first grade. Moreover, no statistically significant differences in total achievement were revealed among the public school comparison groups at the beginning of first grade.

Figure 18. Pairwise Mean Differences in Total Achievement among Student Groups by Testing Episode

Early First Grade (Fall 1998)

3-year Scholarship Recipients	4-year Scholarship Recipients	Applicant Non-Recipients	Non-Applicants	Winner Non-users
		522.61	520.67	520.67
531.94	530.38			

Late First Grade (Spring 1999)

Winner Non-users	3-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	4-year Scholarship Recipients
551.11	549.71	548.02	547.83	546.73

Late Second Grade (Spring 2000)

Winner Non-users	4-year Scholarship Recipients	Non-Applicants	3-year Scholarship Recipients	Applicant Non-Recipients
580.58	578.45	578.30	577.96	577.37

Late Third Grade (Spring 2001)

4-year Scholarship Recipients	Winner Non-users	Non-Applicants	Applicant Non-Recipients	3-year Scholarship Recipients
	605.69	605.51	604.93	600.82
606.46				

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

Additional paired comparisons found no statistically significant differences in total achievement among the groups of students at the end of first grade. Similarly, no statistically significant differences in total achievement were found among the student groups at the end of second grade. By the end of third grade, however, four-year scholarship recipients (M = 606.46) were achieving at a significantly higher level than were three-year scholarship recipients (M = 600.82). No other statistically significant differences in total achievement were found among the student groups at the end of third grade (see Figure 18).

To further examine the nature of the group by testing episode interaction, follow-up analyses examined differences in the amount of change in total achievement that occurred within each student group from one testing episode to another. The results of these comparisons are depicted graphically in Figure 19. From early first grade to late first grade, comparisons revealed that the increase in total achievement for three-year and four-year scholarship recipients was statistically smaller than the increase in total achievement experienced by each of the public school comparison groups. Specifically, three-year scholarship recipients (M = 17.77) and four-year scholarship recipients (M = 16.35) displayed significantly less gain in total achievement than did: applicant non-recipients (M = 25.22), non-applicants (M = 27.35), and winner non-users (M = 30.44). No other statistically significant differences in total achievement gain from the beginning to the end of first grade were found among the student groups.

From the end of first grade to the end of second grade, as well as from the end of second grade to the end of third grade, no statistically significant differences in total achievement gain were found among the groups. Each displayed a similar gain in total achievement from late first grade to late second grade (three-year scholarship recipients, M = 28.25; four-year scholarship recipients, M = 31.72; applicant non-recipients, M = 29.54; non-applicants, M = 30.28; and winner non-users, M = 29.47); and from late second grade to late third grade (three-year scholarship recipients, M = 22.86; four-year scholarship recipients, M = 28.01; applicant non-recipients, M = 27.56; non-applicants, M = 27.21; and winner non-users, M = 25.11).

Figure 19. Pairwise Mean Differences in Total Achievement CHANGE among Student Groups by Testing Episode

Early First Grade (Fall 1998) to Late First Grade (Spring 1999): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	3-year Scholarship Recipients	4-year Scholarship Recipients
			17.77	16.35
30.44	27.35	25.22		

Late First Grade (Spring 1999) to Late Second Grade (Spring 2000): *Scale Score Change*

4-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	Winner Non-users	3-year Scholarship Recipients
31.72	30.28	29.54	29.47	28.25

Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001): *Scale Score Change*

4-year Scholarship Recipients	Applicant Non-Recipients	Non-Applicants	Winner Non-users	3-year Scholarship Recipients
28.01	27.56	27.21	25.11	22.86

Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): *Scale Score Change*

Winner Non-users	Non-Applicants	Applicant Non-Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients
				68.88
			76.08	
85.02	84.84	82.32		

Late First Grade (Spring 1999) to Late Third Grade (Spring 2001): *Scale Score Change*

4-year Scholarship Recipients	Non-Applicants	Applicant Non-Recipients	Winner Non-users	3-year Scholarship Recipients
	57.49	57.10	54.58	51.11
59.73				

Interpretation: The blue underline indicates which means do and do not differ significantly. Specifically, means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level. Further, the means are presented in multiple layers/levels to facilitate the simultaneous display of the pairwise differences among means that are and are not statistically significant at the 0.05 alpha-level.

Overall gain in total achievement from early first grade to late third grade was statistically smaller for three-year and four-year scholarship recipients than it was for any of the public school comparison groups. Each of the public school comparison groups demonstrated significantly greater improvement in total achievement from early first grade to late third grade than did either three-year scholarship recipients (M = 68.88) or four-year scholarship recipients (M = 76.08): applicant non-recipients (M = 82.32), non-applicants (M = 84.84), and winner non-users (M = 85.02). Moreover, four-year scholarship recipients demonstrated a significantly greater improvement in total achievement from early first grade to late third grade than did three-year scholarship recipients, but no statistically significant differences in total achievement gain from early first grade to late third grade were found among the public school comparison groups.

After excluding the initial differences in total achievement at the beginning of first grade, as well as the change that occurred between the beginning and end of first grade, however, the only statistically significant difference in achievement gain was between three-year and four-year scholarship recipients. More specifically, from late first grade to late third grade, four-year scholarship recipients (M = 59.73) displayed a greater gain in total achievement than did three-year scholarship recipients (M = 51.11). No other statistically significant differences in total achievement gain were found among the groups across this time period (applicant non recipients, M = 57.10; non-applicants, M = 57.49; winner non-users, M = 54.58).

Total Achievement Summary. The total achievement score is calculated as the average of reading, language, and mathematics scores for each student. As a result, the pattern of differential achievement across time on the total score was similar to the general pattern displayed on all of the subtests described above. Specifically, significant differences in total achievement that existed between the scholarship students and those in the three public school comparison groups at the beginning of first grade were not present at the end of first grade. All of the public school students made significantly greater gains during this time period in comparison to scholarship recipients, and all of the groups achieved at statistically similar levels by the end of first grade. During second and third grade, however, students in most of the groups made similar gains in total achievement. Although, because of slight differences in achievement gain during third grade, four-year scholarship recipients displayed statistically higher total

achievement scores than did three-year scholarship recipients at the end of third grade, but no other differences among the student groups were statistically significant.

3.3.5 Synopsis across the Four Achievement Measures

Taken together, analyses of the four achievement measures (reading, language, mathematics, and total achievement) revealed three general and relatively consistent findings. (1) At the beginning of first grade, scholarship recipients who had been in the program since kindergarten, as well as scholarship recipients who entered the program as first graders, were achieving at a higher level than their counterparts in public schools. (2) From the beginning to the end of first grade, students in the public schools improved more in terms of academic achievement than did either of the two scholarship recipient groups in private schools. Despite a differential pattern of gain in achievement, however, the public school comparison groups did not gain a sufficient amount on the academic achievement measures to surpass the scholarship recipients. Therefore, at the end of first grade, few differences in achievement were found among all of the student groups. (3) From the end of first grade to the end of second grade, as well as from the end of second grade to the end of third grade, all of the student groups displayed similar patterns of improvement in academic achievement. That is, no one group improved at a consistently or significantly different rate than did any of the other groups. Consequently, at the end of both second grade and third grade, few differences in academic achievement were found among the student groups. Most notably, no consistent differences in achievement between scholarship recipients and public school students were observed at the end of third grade, 2001.

In sum, across all of the measures, a significant initial gap in achievement at the beginning of first grade was closed by public school students by the end of first grade. And, throughout second and third grade, scholarship and public school students continued to improve and perform at statistically similar levels.[54]

[54] With the demographic characteristics removed from the analyses as covariates, independent three-factor mixed-design analyses of variance failed to reveal statistically significant three-way interactions among: (a) minority status, student group, and testing episode, (b) sex, student group, and testing episode, and (c) estimated meal code, student group, and testing episode. That is, none of the analyses found that the student group by testing episode interaction effects (discussed above) differed as a function of minority status, sex, or meal code status (as a proxy for family income). Although the pattern of change in achievement across time varied depending on the group to which students belonged (e.g., scholarship recipients vs. public non-applicants), this differential pattern of change among the student groups was not statistically different for: (a) minority students versus non-minority students, (b) males students versus female students, or (c) students who received free or reduced lunches versus students who paid for their lunches.

3.3.6 Effect Size Analysis: Magnitude and Practical Significance of Achievement Score Interaction Effects

As discussed above, the student group by testing episode interaction was statistically significant across all of the achievement tests, indicating that various groups of students demonstrated different patterns of change in achievement across time. Effect size analyses, however, suggest that the magnitude of the differences among patterns of achievement change displayed by the groups of students is small. This section addresses the practical significance of the interactions found on the reading, language, mathematics, and total scores. Eta-squared statistics were calculated for each effect of interest in the factorial mixed-design analyses of covariance discussed above. Eta is a measure of association that is appropriate for a categorical independent variable(s) and a continuous dependent variable. Eta squared can be interpreted as the proportion of variance in the dependent variable (i.e., achievement scale scores) explained by differences among student groups—thus, eta squared can range from .00 to 1.0.

Tables 24, 26, 28, and 30 (above) display the p-values and effect size statistics (eta squared) associated with the main effects and interaction for each achievement measure. Across all of the achievement scores (reading, language, mathematics, and total), the main effect of testing episode (time) accounted for the largest portion of variance in academic achievement—ranging from 3% in language achievement to approximately 16% in Reading achievement. That is, the progressive increase in achievement scores from one year to the next, regardless of student group membership, was the largest effect in each of the analyses discusses above. Moreover, collapsing across the testing episodes, no statistically significant differences were found in achievement among the student groups (i.e., no main effect of student group membership). Therefore, it is not surprising that small differences among the student groups, after averaging across the four testing episodes, account for considerably less than 1% of the variance in each of the achievement scores.

Most notably, the group by testing episode interaction effects, although statistically significant, explain only a small proportion of the variance in academic achievement. That is, the magnitude of the interaction effect on each achievement measure is relatively small, accounting for less than 0.1% to only about 2% of the variance in academic achievement (see Tables 24, 26, 28, and 30). Although the interaction effects indicate that the pattern of achievement change across time was

not the same for all student groups, the size of the differences in change was not large (as indicated by small eta squared statistics). In other words, small interaction effect sizes indicate that the pattern of achievement from first grade through third grade varies only slightly as a function of group membership. Therefore, some caution should be exercised when interpreting these statistically significant interaction effects because, from a practical standpoint, the change in academic achievement across time was relatively similar for all of the student groups.

3.3.7 Differential Entry into the CSTP and Academic Achievement

Another question regarding the impact of the CSTP on academic achievement is whether the amount of time students have spent in the program is related to achievement. Specifically, do scholarship recipients who have spent more time in the program achieve at a higher level, or experience greater achievement gains across time, in comparison to recipients who have spent less time in the program? In other words, is the impact of the CSTP cumulative, requiring several years of participation before any benefits become manifest in achievement scores, or is the duration of participation unrelated to academic achievement? Four groups of scholarship recipients were identified for the following analyses based on when they entered the program. Specifically, *four-year scholarship recipients* entered program as kindergartners in autumn, 1997; *three-year scholarship recipients* entered the program as first graders in autumn, 1998; *two-year scholarship recipients* entered the program as second graders in autumn, 1999; and *one-year scholarship recipients* entered the program as third graders in autumn, 2000.[55]

Correlational Analysis of Duration of Participation and Achievement

Correlations between the amount of time that students have participated in the program (ranging from one year to four years) and academic achievement at end of third grade (spring, 2001) indicate that the duration of participation in the program is unrelated to third grade achievement. This means that, among scholarship recipients, the high achievers in third grade are as likely to have participated in the program since kindergarten (for four years) as they are to have

[55] When it was possible to identify a student's school of prior enrollment, scholarship recipients from private schools were *not* included in these analyses. That is, an attempt was made to restrict the analyses of achievement differences based on differential entry into the CSTP to students who attended *public schools* prior to receiving and using a scholarship. As a result, 62 out of 110 one-year scholarship recipients and 47 out of 79 two-year scholarship recipients were identified for the analyses as either (a) students from public school or (b) students for whom their prior school of enrollment was not known. Inclusion of these students in the achievement score analyses also was conditional upon the availability of their achievement scores.

participated only since the beginning of third grade (for one year). There is no consistent linear pattern of higher achievement associated with more time in the program.

Correlational analyses also examined the relationship between the amount of time recipients have spent in the CSTP and the magnitude of achievement gain from the end of second grade to the end of third grade. By the end of second grade, three groups of scholarship recipients were participating in the program: students who first entered in kindergarten, first grade, and second grade. By the end of third grade, the fourth group had participated in the program for one year. Therefore, achievement gain from second to third grade was identified as the most appropriate measure to correlate with the duration of participation in the program.

Unlike the analyses of third grade achievement, the duration of participation in the program was positively correlated with achievement gain from second to third grade on all of the achievement measures except language. In other words, recipients who had spent more time in the program tended to gain more over this time period in reading, mathematics, and total achievement in comparison to recipients who had spent less time in the program. Table 31 below displays these correlations. However, the amount of variance in achievement gain explained by the duration of participation in the CSTP was relatively small (approximately 1%). Furthermore, because of marked differences in the size of the four scholarship recipient groups, the largest groups (four-year and three-year recipients) received more weight in the correlational analysis than did the smallest groups (two-year and one-year recipients). Thus, the positive correlations likely reflect that four-year recipients tended to gain more on the achievement measures from second to third grade than did three-year recipients. To further examine the nature of achievement and achievement gain as a function of differential entry into the CSTP, a series of analyses of variance were conducted and the results are discussed below.

Table 31. Correlations between Duration of Participation in the CSTP and Achievement Score Gain from Second to Third Grade

	Correlation Coefficients		
	Reading	Mathematics	Total
Correlation between (1) achievement gain from late 2nd grade to late 3rd grade and (2) duration of participation in the CSTP	**.091**	**.115**	**.105**
p-value	.046	.012	.021
N	480	480	480

Group Comparisons: Third Grade Achievement (Spring, 2001) as a Function of Differential Entry

To further examine whether and how the academic achievement of students who entered the CSTP later differs from the achievement of students who entered earlier (e.g., as third graders rather than as first graders), analyses of variance were conducted comparing the spring, 2001 achievement scores of students who entered the CSTP as: (a) kindergartners in autumn, 1997; (b) first graders in autumn, 1998; (c) second graders in autumn, 1999; and (d) third graders in autumn, 2000. Additional analyses were conducted to examine the differences in achievement between the one-year and two-year scholarship recipients and the public school comparison groups within the late third grade testing episode.[56] To avoid excluding cases from the relatively small one-year and two-year scholarship recipient groups, the student demographic variables of sex, minority status, and estimated meal code were not included in the analyses as covariates. Table 32 displays the reading, language, mathematics, science, social studies, and total achievement scale score means from the most recent testing episode (late third grade, 2001) for each scholarship recipient group.

The level of third grade achievement in reading, language, and social studies was not statistically different among the scholarship recipient groups as a function of differential entry into the program. Specifically, no statistically significant mean differences were found among the recipient groups in 2000-2001 *reading* ($F[3, 571] = 1.90$, $p = .13$), *language* ($F[3, 571] = 1.19$, $p = .31$), or *social studies* ($F[3, 519] = 2.11$, $p = .10$). In contrast, statistically significant differences in achievement were found among the differential entry groups on the *mathematics* ($F[3, 570] = 3.11$, $p = .03$), *total* ($F[3, 568] = 2.65$, $p = .05$), and *science* measures ($F[3, 519] = 3.21$, $p = .02$). However, after controlling for multiple comparisons, follow-up analyses revealed only one statistically significant mean difference among the recipient groups. Specifically, four-year scholarship recipients achieved at a significantly higher level on the science achievement measure ($M = 595.21$) in comparison to two-year scholarship recipients ($M = 577.03$). No other statistically significant mean differences on the mathematics, total, or science achievement measures were found between any of the recipient groups after controlling for multiple comparisons. Furthermore, orthogonal comparisons between early entrants into the program

[56] Comparisons among three-year recipients, four-year recipients, and the public school groups are presented and discussed in the preceding section of this report.

(three- and four-year recipients) and recent entrants (one- and two-year recipients) on each of the measures failed to reveal statistically significant differences in mathematics, total, or science achievement.

Table 32. Scholarship Student Achievement (Late Third Grade, 2001) Based on Differential Entry into the Program

		Reading		Language		Mathematics		Total Achievement	
Scholarship Status	N	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1-year Scholarship Recipients	54	**618.57**	5.17	**613.24**	4.38	**595.00**	5.05	**609.83**	4.42
2-year Scholarship Recipients	41	**607.39**	5.93	**605.15**	5.03	**581.02**	5.74	**597.08**	5.04
3-year Scholarship Recipients	119	**611.00**	3.48	**612.10**	2.95	**585.23**	3.37	**602.79**	2.92
4-year Scholarship Recipients	361	**618.26**	2.00	**614.81**	1.70	**594.27**	1.93	**609.11**	1.68
Public Applicant Non-Recipients	419	**613.04**	1.87	**607.50**	1.66	**590.01**	1.93	**603.52**	1.62
Public Winner Non-Users	85	**613.09**	4.16	**604.95**	3.68	**591.05**	4.27	**603.03**	3.60
Public Non-Applicants	513	**612.75**	1.69	**608.42**	1.50	**593.31**	1.74	**604.81**	1.47

Total Achievement = Average of Reading, Language, and Mathematics Measures

		Science		Social Studies	
Scholarship Status	N	Mean	SE	Mean	SE
1-year Scholarship Recipients	54	**589.26**	5.48	**608.80**	4.76
2-year Scholarship Recipients	40	**577.03**	6.37	**596.40**	5.53
3-year Scholarship Recipients	101	**586.57**	4.01	**606.34**	3.48
4-year Scholarship Recipients	328	**595.21**	2.22	**610.59**	1.93
Public Applicant Non-Recipients	337	**591.18**	2.77	**606.25**	2.16
Public Winner Non-Users	71	**586.00**	6.03	**607.25**	4.70
Public Non-Applicants	482	**590.59**	2.31	**607.62**	1.80

Mathematics Measure: 1-year Scholarship Recipients N = 53.
Total Achievement Measure: 1-year Scholarship Recipients N = 52, 2-year Scholarship Recipients N = 40.

In comparison to the public school groups, scholarship recipients who entered the program as third graders descriptively achieved at a slightly higher level on all of the measures except science at the end of third grade. Despite this trend, however, none of the differences in 2000-

2001 achievement between the differential entry recipient groups and the public school groups were statistically significant: *reading*: (F[4, 1107] = .509 , p = .73): *language* (F[4, 1107] = .608, p = .66), *mathematics* (F[4, 1106] = 1.26, p = .29), *total* (F[4, 1104] = .948, p = .44), *social studies* (F[4, 979] = .794, p = .53), and *science* (F[4, 979] = .824, p = .51). This means that the observed differences between the groups are not meaningful and the groups should be considered equivalent.

Prior analyses conducted on achievement scores from each of the previous testing episodes also failed to find statistically significant differences among the scholarship recipient groups based on when they entered the program.[57] The present findings are consistent with previous results reported elsewhere and, taken together, indicate that former public school students who enter the CSTP do not initially achieve at a statistically different level on any of the tests in comparison to either (a) scholarship recipients in the program prior to a given testing episode under consideration, or (b) their peers in public schools. That is, achievement differences have not been found based on the duration of students' participation in the CSTP.

Group Comparisons: Achievement Gain from Second Grade to Third Grade as a Function of Differential Entry

Analyses of achievement *change* (gain) across time also were conducted to examine whether students who entered the program earlier demonstrated greater or lesser levels of achievement gain in comparison to students who entered the program more recently. Complete reading, language, mathematics, and total achievement scale scores from first through third grade were available for only 7 of the 62 public school students who entered the CSTP as third graders in autumn, 2000, and only 10 of the 47 public school students who entered the program as second graders in autumn, 1999.[58] Furthermore, because of the relatively small total number of students and the large proportion of missing data present within these groups in the first and second grade testing episodes, multiple imputation procedures were not performed to estimate and replace missing values. Thus, differences in achievement change were analyzed only from second to third grade, which provided a greater number of complete cases within each scholarship recipient

[57] For details of these analyses, see Metcalf, K.K. *Cleveland Scholarship and Tutoring Program Evaluation: 1998-2000 Technical Report.* (Bloomington, Indiana; Indiana Center for Evaluation, 2001).

[58] Science and Social Studies scores could not be analyzed because these subtests were administered for the first time in third grade (spring, 2001).

group.[59] To avoid excluding cases from the small one-year and two-year scholarship recipient groups, the student demographic variables of sex, minority status, and estimated meal code were not included in the analyses as covariates.

In separate mixed-design analyses of variance incorporating one between-subjects variable (student group) and one within-subject variable (time of testing), late second grade to late third grade gain in reading, language, mathematics, and total achievement for one-year and two-year scholarship recipients was compared to: (a) scholarship recipients who entered the program earlier, either in kindergarten or in first grade, and (b) the three public school comparison groups (applicant non-recipients, non-applicants, and winner non-users). Notably, for scholarship recipients who entered the program as third graders, the following achievement change analyses offer the opportunity to examine achievement before and after participation in the CSTP, and whether this change differs from other scholarship recipient or public school comparison groups.

Reading Achievement Change among Scholarship Recipients: Descriptive statistics for the analysis of reading achievement change across time, which include the four scholarship recipient groups based on differential entry into the program, are presented in Table 33. Figure 20 presents these data graphically. Results of the mixed-design analysis of variance are displayed in Table 34.[60]

The analysis of reading achievement found a statistically significant main effect of time (i.e., testing episode). The main effect indicates that, regardless of when students entered the CSTP, all scholarship recipients improved significantly in reading achievement from late second grade to late third grade. That is, independent of the point at which they entered the program, student reading achievement was statistically higher at the end of third grade than it was at the end of

[59] Because the number of scholarship recipients who provided a complete set of achievement scores in reading, language, and mathematics both in the second grade and in the third grade differs from the number of students in the analyses of only third grade achievement (reported above), discrepancies in achievement differences have resulted. The following is the number of one-year and two-year scholarship recipients who provided two consecutive years of achievement data from second to third grade on each test: *reading*, 16 one-year and 36 two-year recipients; *language*, 16 one-year and 36 two-year recipients; *mathematics*, 15 one-year and 36 two-year recipients, and *total* 14 one-year and 35 two year recipients (i.e., they had testing episode). Because the number of students in each of these groups is small, caution should be exercised when interpreting the results of these analyses.

[60] The assumption of sphericity (i.e., homogeneity applied to variance of the differences between levels of the within-subjects variable) cannot be violated when the within-subject variable has only two levels. It is impossible to have heterogeneous variance of the differences between levels of the within-subjects variable with only two levels.

second grade. The main effect of scholarship recipient group, however, was not statistically significant. After averaging across the two testing periods, reading achievement did not differ significantly among the scholarship recipient groups.

Table 33. Student Achievement in Reading: Late Second Grade (Spring, 2000) to Late Third Grade (Spring 2001) Based on Differential Entry into the CSTP

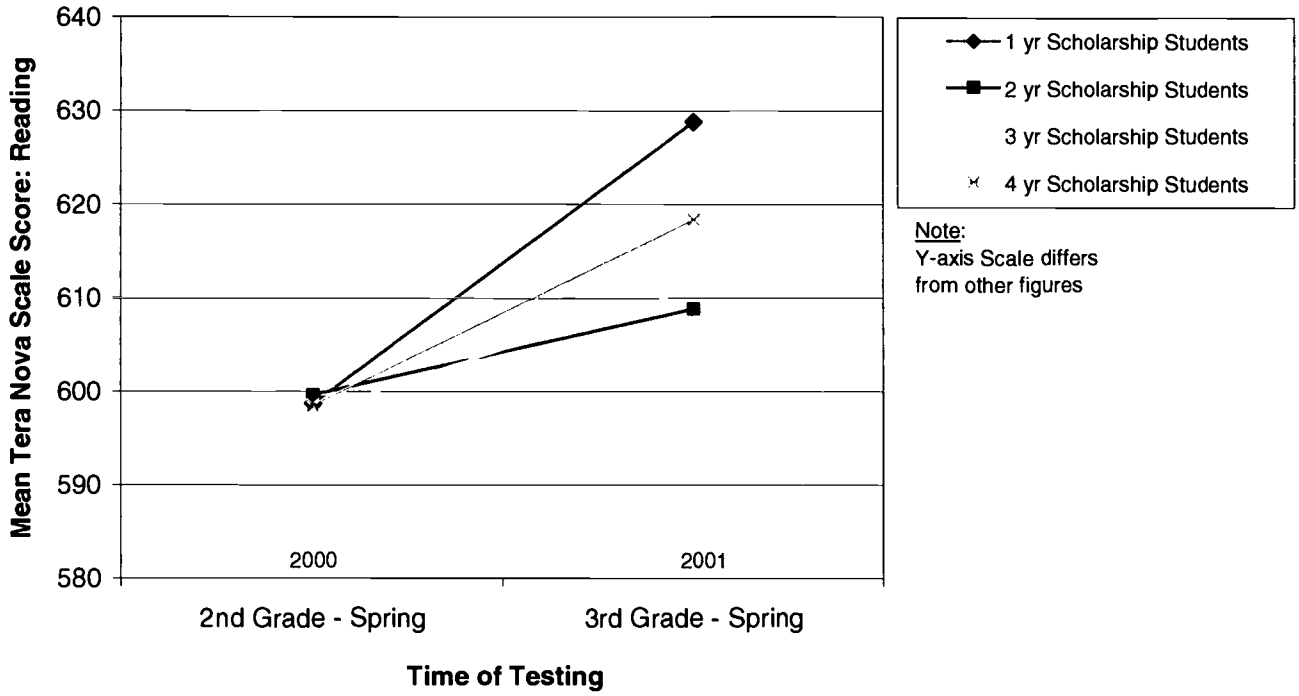
Student group	N	Spring 2nd Grade (2000) Mean	SE	Spring 3rd Grade (2001) Mean	SE	Marginal Means (Student group) Mean	SE
1-year Scholarship Recipients	16	**598.68**	8.54	**628.75**	9.53	**613.72**	8.15
2-year Scholarship Recipients	36	**599.47**	5.69	**608.64**	6.35	**604.06**	5.43
3-year Scholarship Recipients	119	**597.80**	3.13	**611.00**	3.50	**604.40**	2.99
4-year Scholarship Recipients	361	**598.42**	1.80	**618.26**	2.01	**608.34**	1.72
Marginal Means (Testing Episode)		**598.60**	2.72	**616.66**	3.03		

Table 34. Mixed Design Analysis of Variance on Reading Achievement Scale Scores: Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001) based on Differential CSTP Entry

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	25745.27	1	25745.27	51.815	< .001	.080
Time x Group	4552.99	3	1517.67	3.054	.028	.010
Error	262347.52	528	496.87			
Between Subjects						
Group	4854.20	3	1618.07	.762	.516	< .001
Error	1121074.43	528	2123.24			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 20. Reading Achievement of Scholarship Recipients Based on Differential Entry: 2000 to 2001



However, a statistically significant interaction was present, indicating that the pattern of change in reading performance differed between at least two of the scholarship recipient groups. Follow-up comparisons were performed to further examine how reading achievement gain varied among the groups from second to third grade. No statistically significant differences in reading achievement were found among the scholarship recipients at the end of second grade. By the end of third grade, however, four-year scholarship students who entered the program as kindergartners were achieving at a statistically higher level in reading (M = 628.75) in comparison both to two-year and to three-year scholarship recipients (M = 608.63 and M = 611.00, respectively). Similarly, one-year scholarship recipients who entered the program as third graders were achieving at a statistically higher level in reading (M = 618.26) in comparison both to two-year and to three-year scholarship recipients. However, at the end of third grade, one-year and four-year recipients did not differ significantly in reading achievement, and two-

year and three-year recipients did not differ significantly in reading achievement. The results of these comparisons are depicted graphically in Figure 21. [61]

Figure 21. Pairwise Differences in Reading Achievement among Differential Entry Scholarship Recipient Groups by Testing Episode

Late Second Grade (Spring 2000): *Mean Scale Score*

3-year Scholarship Recipients	4-year Scholarship Recipients	1-year Scholarship Recipients	2-year Scholarship Recipients
599.47	598.69	598.42	597.80

Late Third Grade (Spring 2001): *Mean Scale Score*

1-year Scholarship Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients	2-year Scholarship Recipients
		611.00	608.64
628.75	618.26		

Interpretation: The blue underline indicates which means do and do not differ significantly. Means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level.

Follow-up comparisons of achievement score change revealed that one-year scholarship recipients, who entered the program in third grade, experienced a significantly greater gain in reading achievement from the end of second grade to the end of third grade (M = 30.06) than did two-year scholarship recipients who entered the program in second grade (M = 9.17). However, the second largest descriptive gain in reading achievement, demonstrated by four-year scholarship recipients (M = 19.85), was not statistically different from the gains experienced by the other scholarship recipient groups: one-year recipients (M = 30.06), two-year recipients (M = 9.17), three-year scholarship recipients (M = 13.20). No other statistically significant differences in reading achievement gain from second to third grade were found among the groups of scholarship recipients. The results of these comparisons are depicted graphically in Figure 22.

[61] The third grade achievement differences reported in this section were not statistically significant when only the third grade data were analyzed using a slightly larger sample of scholarship recipients (see the preceding section). Therefore, it is difficult to meaningfully interpret these results.

Figure 22. Pairwise Differences in Reading Achievement CHANGE among Differential Entry Scholarship Recipient Groups by Testing Episode

Late Second Grade (Spring 2000) to Late Third Grade: *Scale Score Change*

1-year Scholarship Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients	2-year Scholarship Recipients
	19.86	13.20	9.16
30.06			

Interpretation: The blue underline indicates which means do and do not differ significantly. Means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level.

Language Achievement Change among Scholarship Recipients: Descriptive statistics for the analysis of language achievement change across time, which include the four scholarship recipient groups based on differential entry into the program, are presented in Table 35. Figure 23 presents these data graphically. Results of the mixed-design analyses of variance on language achievement are displayed in Table 36.

Table 35. Student Achievement in Language: Late Second Grade (Spring, 2000) to Late Third Grade (Spring 2001) Based on Differential Entry into the CSTP

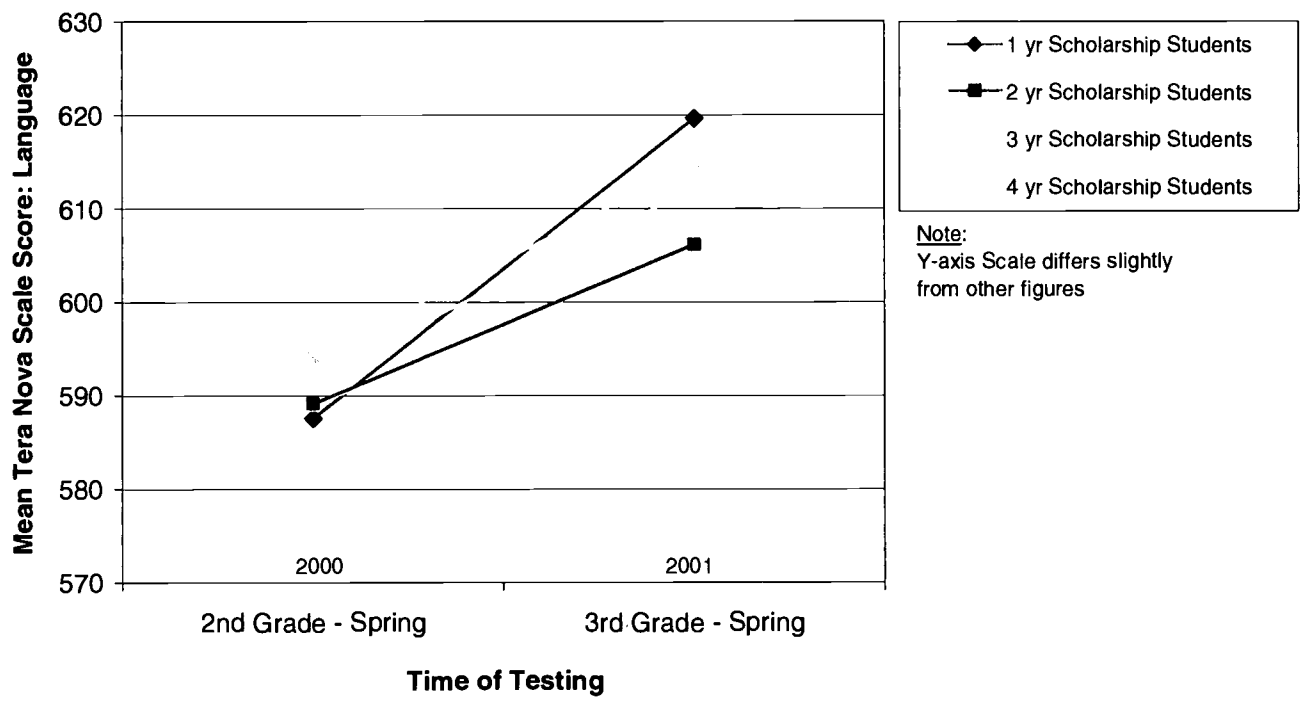
Student Group	N	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student Group) Mean	Marginal Means (Student Group) SE
1-year Scholarship Recipients	16	**587.50**	9.05	**619.56**	8.06	**603.53**	7.67
2-year Scholarship Recipients	36	**588.97**	6.03	**605.97**	5.37	**597.47**	5.11
3-year Scholarship Recipients	119	**592.04**	3.32	**612.10**	2.95	**602.07**	2.81
4-year Scholarship Recipients	361	**594.01**	1.91	**614.81**	1.70	**604.41**	1.62
Marginal Means (Testing Episode)		**590.63**	2.88	**613.11**	2.56		

Table 36. Mixed Design Analysis of Variance on Language Achievement Scale Scores: Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001) based on Differential CSTP Entry

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	39850.37	1	39850.37	85.443	<.001	.130
Time x Group	1305.91	3	435.30	.933	.424	<.001
Error	246257.35	528	466.39			
Between Subjects						
Group	3685.03	3	1228.34	.653	.582	<.001
Error	993858.04	528	1882.30			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 23. Language Achievement of Scholarship Recipients Based on Differential Entry: 2000 to 2001



For language achievement, the analysis found a statistically significant main effect of time (i.e., testing episode) indicating that, regardless of when students entered the CSTP, all scholarship recipients improved significantly in reading achievement from late second grade to late third grade. The main effect of scholarship group, however, was not statistically significant. After averaging across the two testing periods, language achievement did not differ significantly among the scholarship recipient groups. Most notably, the interaction between the scholarship recipient groups and testing episode was not statistically significant with respect to language achievement. That is, the pattern of language achievement change from the end of second grade to the end of third grade did not differ depending on when students entered the CSTP (i.e., all of the scholarship recipient groups displayed statistically similar patterns of achievement across the two testing episodes).

Mathematics Achievement Change among Scholarship Recipients: Descriptive statistics for the analysis of mathematics achievement change across time, which include the four scholarship recipient groups based on differential entry into the program, are presented in Table 37. Figure 24 presents these data graphically. Results of the mixed-design analyses of covariance on mathematics achievement are displayed in Table 38.

The analysis of mathematics achievement found a statistically significant main effect of time (i.e., testing episode) indicating that, regardless of when they entered the CSTP, all scholarship recipients improved significantly in mathematics achievement from late second grade to late third grade. That is, independent of the point at which they entered the program, student achievement in mathematics was statistically higher at the end of third grade than is was at the end of second grade. The main effect of scholarship recipient group membership, however, was not statistically significant. After averaging across the two testing periods, mathematics achievement did not differ significantly among the scholarship recipient groups. Furthermore, the student group by testing episode interaction was not statistically significant, indicating that the pattern of mathematics achievement gain from the end of second grade to the end of third grade did not differ as a function of when students entered the CSTP. In other words, all of the scholarship recipient groups displayed statistically similar patterns of mathematics achievement gain across the two testing episodes.

Table 37. Student Achievement in Mathematics: Late Second Grade (Spring, 2000) to Late Third Grade (Spring 2001) Based on Differential Entry into the CSTP

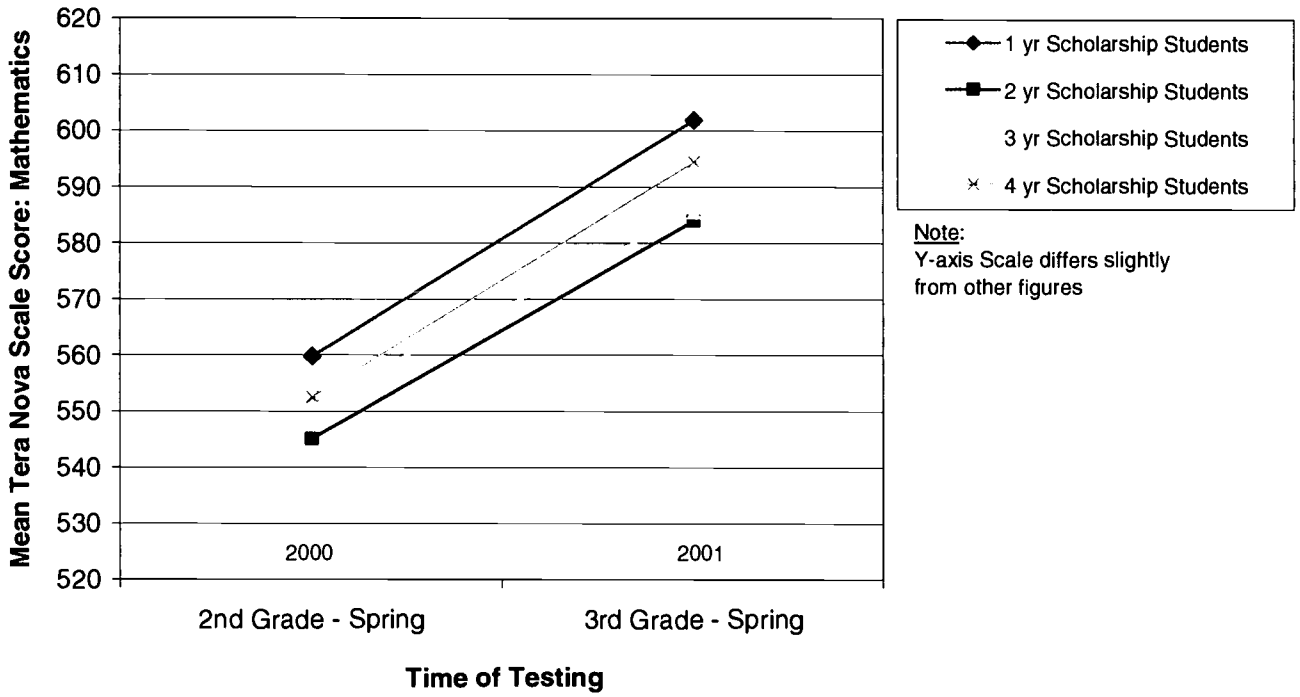
Student Group	N	Spring 2nd Grade (2000) Mean	SE	Spring 3rd Grade (2001) Mean	SE	Marginal Means (Student Group) Mean	SE
1-year Scholarship Recipients	15	**559.73**	8.69	**601.80**	9.45	**580.77**	8.15
2-year Scholarship Recipients	36	**544.86**	5.61	**583.67**	6.10	**564.26**	5.26
3-year Scholarship Recipients	119	**551.61**	3.09	**585.23**	3.36	**568.42**	2.89
4-year Scholarship Recipients	361	**552.38**	1.77	**594.27**	1.93	**573.33**	1.66
Marginal Means (Testing Episode)		**552.14**	2.73	**591.24**	2.97		

Table 38. Mixed Design Analysis of Variance on Mathematics Achievement Scale Scores: Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001) based on Differential CSTP Entry

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	115794.35	1	115794.35	241.076	<.001	.310
Time x Group	3114.95	3	1038.32	2.162	.092	.010
Error	253130.73	527	480.32			
Between Subjects						
Group	10902.08	3	3634.03	1.824	.142	.010
Error	1050140.18	527	1992.67			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 24. Mathematics Achievement of Scholarship Recipients Based on Differential Entry: 2000 to 2001



Total Achievement Change among Scholarship Recipients: Descriptive statistics for the analysis of total achievement change across time, which included the four scholarship recipient groups based on differential entry into the program, are presented in Table 39. Figure 25 presents these data graphically. Results of the mixed-design analyses of variance on total achievement are displayed in Table 40. A statistically significant main effect of time (i.e., testing episode) was found, indicating that, regardless of when students entered the CSTP, all scholarship recipients improved significantly in total achievement from late second grade to late third grade. That is, independent of the point at which they entered the program, student achievement was statistically higher at the end of third grade than it was at the end of second grade. The main effect of scholarship recipient group membership, however, was not statistically significant. After averaging across the two testing periods, total achievement did not differ significantly among the scholarship recipient groups.

Table 39. Student Total Achievement: Late Second Grade (Spring, 2000) to Late Third Grade (Spring 2001) Based on Differential Entry into the CSTP

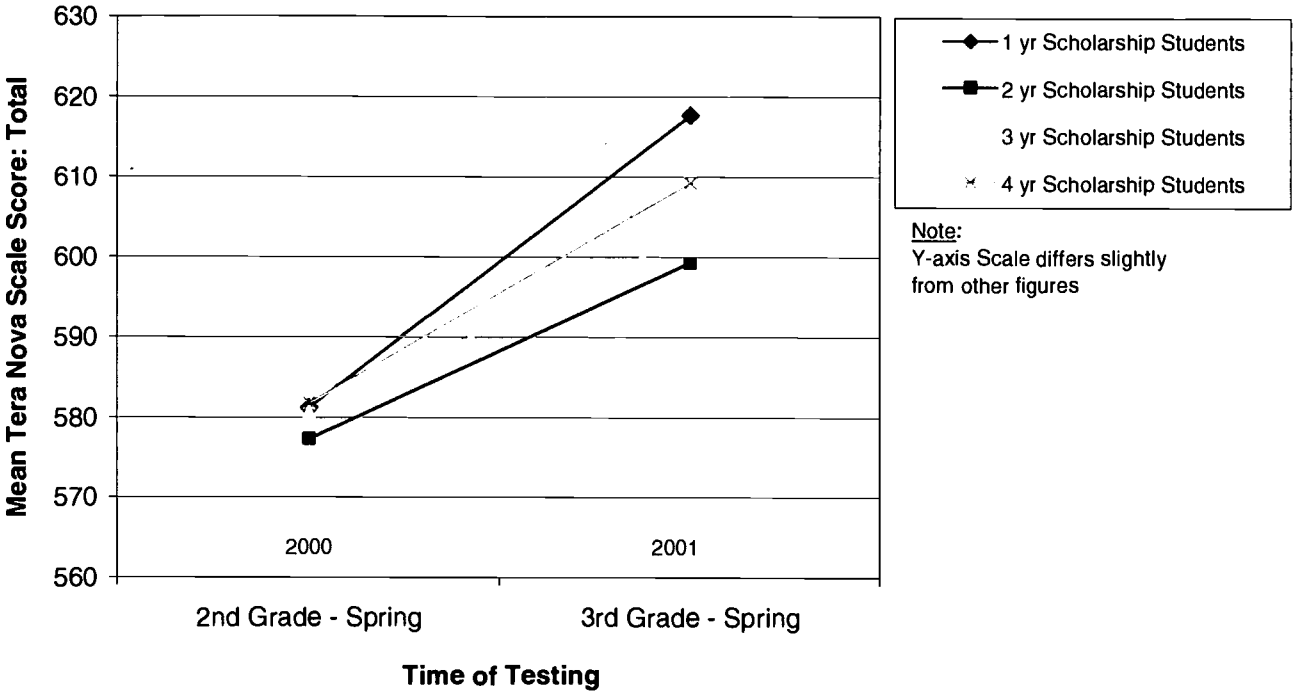
Student Group	N	Spring 2nd Grade (2000) Mean	SE	Spring 3rd Grade (2001) Mean	SE	Marginal Means (Student Group) Mean	SE
1-year Scholarship Recipients	14	**581.14**	8.02	**617.64**	8.54	**599.39**	7.79
2-year Scholarship Recipients	35	**577.14**	5.07	**599.11**	5.40	**588.13**	4.93
3-year Scholarship Recipients	119	**580.59**	2.75	**602.79**	2.93	**591.69**	2.67
4-year Scholarship Recipients	361	**581.72**	1.58	**609.11**	1.68	**595.42**	1.53
Marginal Means (Testing Episode)		**580.15**	2.50	**607.17**	2.66		

Table 40. Mixed Design Analysis of Variance on Total Achievement Scale Scores: Late Second Grade (Spring 2000) to Late Third Grade (Spring 2001) based on Differential CSTP Entry

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	52516.35	1	52516.35	233.767	<.001	.300
Time x Group	2264.54	3	754.85	3.360	.019	.010
Error	117942.37	525	224.65			
Between Subjects						
Group	5902.90	3	1967.64	1.159	.325	<.001
Error	891335.34	525	1697.78			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 25. Total Achievement (Average Reading, Language, and Mathematics Achievement) of Scholarship Recipients Based on Differential Entry: 2000 to 2001



The scholarship recipient group by testing episode interaction, however, was statistically significant, indicating that the pattern of total achievement gain from late second grade to late third grade differed between at least two of the scholarship recipient groups. To examine the nature of the interaction, paired comparisons were performed. No statistically significant differences in total achievement were found among the scholarship recipient groups at the end of second grade. By the end of third grade, however, one-year recipients (M = 617.64) and four-year recipients (M = 609.11) were achieving at a statistically higher level than were two-year recipients (M = 599.11) and three-year recipients (M = 602.79). No other statistically significant differences in total academic achievement were found among the groups at the end of third grade. The results of these comparisons are depicted graphically in Figure 26.[62]

[62] The third grade achievement differences reported in this section were not found to be statistically significant when only the third grade data were analyzed using a slightly larger sample of scholarship recipients (see the preceding section). Therefore, it is difficult to interpret these results.

Figure 26. Pairwise Differences in Total Achievement among Differential Entry Scholarship Recipient Groups by Testing Episode

Late Second Grade (Spring 2000): *Mean Scale Score*

4-year Scholarship Recipients	1-year Scholarship Recipients	3-year Scholarship Recipients	2-year Scholarship Recipients
581.72	581.14	580.59	577.14

Late Third Grade (Spring 2001): *Mean Scale Score*

1-year Scholarship Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients	2-year Scholarship Recipients
		602.79	599.11
617.64	609.11		

Interpretation: The blue underline indicates which means do and do not differ significantly. Means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level.

Furthermore, follow-up analyses of the change in total achievement displayed by each scholarship recipient group from the end of second grade to the end of third grade revealed that one-year scholarship students who entered the program as third graders made greater gains in achievement (M = 36.50) than did three-year scholarship recipients who participated in the program since first grade (M = 22.19). That is, after participating in the CSTP for one academic year, the most recent scholarship recipient group demonstrated greater improvement in total achievement from the end of second grade (before these students participated in the program) to the end of third grade in comparison to scholarship recipients who had been in the program for three academic years (first grade through third grade). Similarly, four-year scholarship students who entered the program as kindergartners made greater gains in achievement (M = 27.39) than did three-year scholarship recipients who participated in the program since first grade. No other differences in total achievement gain among the scholarship recipient groups were statistically significant. The results of these comparisons are depicted graphically in Figure 27.

Figure 27. Pairwise Differences in Total Achievement CHANGE among Differential Entry Scholarship Recipient Groups by Testing Episode

Late Second Grade (Spring 2000) to Late Third Grade: *Scale Score Change*

1-year Scholarship Recipients	4-year Scholarship Recipients	3-year Scholarship Recipients	2-year Scholarship Recipients
		22.19	21.97
36.50	27.39		

Interpretation: The blue underline indicates which means do and do not differ significantly. Means that are connected by a common underlined DO NOT differ at a statistically significant level. Similarly, if the blue underline associated with a mean or group of means OVERLAPS with another mean or group of means, this indicates that none of those means differ at a statistically significant level.

Group Comparisons: Achievement Gain of One-Year and Two-Year Scholarship Recipients versus Public School Comparison Groups

Mixed-design analyses of variance on reading, language, mathematics, and total achievement scores were conducted to examine whether the achievement gain of recent entrants into the CSTP differed from that of the public school comparison groups.[63] One-year scholarship recipients, two-year scholarship recipients, and the three public school comparison groups were included as the levels of the between-subjects factor (group membership). Testing episode (late second grade and late third grade) served as the within-subjects factor.

On each of the measures of academic achievement (reading, language, mathematics, and total scale scores), the analysis revealed a statistically significant main effect of time (i.e., testing episode) indicating that, regardless of group membership, achievement was statistically higher at the end of third grade than is was at the end of second grade. The main effect of group membership, however, was not statistically significant on any of the measures of achievement. In other words, after averaging across the two testing episodes, achievement on each measure did not differ significantly among the groups. Furthermore, the student group by testing episode interaction was not statistically significant, indicating that the pattern of achievement change (gain) from the end of second grade to the end of third grade did not differ depending on group membership. In other words, the scholarship recipient groups and their counterparts in the public

[63] Analyses comparing the achievement gain of four-year and three-year scholarship recipients to the public school comparison groups were conducted and reported in a previous section.

school comparison groups displayed statistically similar patterns of achievement gain across the two testing episodes.

Summary: Differential Entry and Academic Achievement

Limitation: Two sets of analyses examined third grade achievement differences among the scholarship recipient groups that entered the program from kindergarten to third grade. The first set examined all recipients who provided complete third grade achievement data. The second set of analyses examined achievement gain from second to third grade, which also included an analysis of third grade achievement differences among the groups. However, the latter analysis was restricted to only recipients who provided complete achievement data both in second and in third grade. Consequently, a smaller sample was used to examine achievement gain as a function of differential entry. Because of relatively small sample sizes in both sets of analyses, and because the sample sizes differ, discrepancies in the statistical significance of the findings emerged. Therefore, interpretation of the differences reported above and summarized below is not only difficult, but also should be done with caution.

Third Grade Achievement: At the end of third grade, the duration of participation in the CSTP was not related to achievement scores. Furthermore, no differences in reading, language, mathematics, social studies, or total achievement scores were found among any of the scholarship recipient groups (based on the first set of analyses utilizing larger sample sizes). The only significant difference found at the end of third grade was in science achievement. Scholarship recipients who participated in the program since kindergarten achieved at a significantly higher level on the third grade science measure in comparison to recipients who entered the program as second graders.

In contrast to the first set of analyses, however, significant differences in third grade reading and total achievement were found using a smaller sample of students in the analysis of achievement gain from second to third grade. In the achievement gain analysis, both four-year recipients, who used scholarships since kindergarten, and one-year recipients, who used scholarships for the first time in third grade, achieved at a higher level on the reading and total measure in comparison to three-year and two-year recipients who entered the program in first and second grade, respectively. Regardless of the analysis employed, however, no differences in third grade

academic achievement were found between any of the scholarship recipient groups and the public school comparison groups.

Achievement Gain: Similarly, few differences were found in the amount of achievement gain displayed from second to third grade. Most notably, one-year scholarship recipients who entered the program in third grade (2000-2001) demonstrated a statistically greater gain from second to third grade on the reading achievement measure in comparison to two-year recipients who entered the program as second graders, and on the total achievement measure in comparison to three-year recipients who entered the program as first graders. That is, after one-year of participation in the CSTP, the most recent recipients displayed a greater gain in reading from second to third grade than did recipients who participated in the program for two years, as well as a greater gain in total achievement than did recipients who participated for three years. Furthermore, four-year scholarship recipients who participated in the program since kindergarten displayed a greater gain in total achievement than did three-year recipients who participated since first grade. The first entrants and the most recent entrants to the program, however, did not differ in the amount of achievement gain from second to third grade.

Taken together, the analyses of differential entry and achievement reveal that neither third grade achievement nor achievement change from second to third grade differ systematically as function of the amount of time recipients have spent in the program. Few differences exist among the scholarship recipient groups, and the differences that were found indicate that recipients who have spent both the most and the least amount of time in the program (four years and one year, respectively) achieved and gained at higher levels than recipients who have spent intermediate amounts of time in the program (two and three years). Moreover, no differences in third grade achievement or in recent achievement gain were found between the scholarship recipient groups and their peers in the public school groups. Although the correlational analysis suggests that the amount of time recipients have participated in the program is positively related to achievement gain from second to third grade, the relationship is weak and the other analyses indicate that more time in the program is not consistently associated with achievement or achievement gain.

3.3.8 Differential Exit from the CSTP: Former Scholarship Recipients and Academic Achievement.

Analyses were conducted to examine whether the academic achievement of former scholarship recipients, who used the scholarship to attend private schools for one or more years and then exited the program and enrolled in public schools, differed from the academic achievement of scholarship recipients who remained in the program. Three groups of *former* scholarship recipients were identified and included in the following analyses: (a) 38 students who exited from the program after one year of participation (kindergarten only) and have attended public schools for three years, from first grade through third grade (*3-year former scholarship recipients*); (b) 40 students who exited from the program after two years of participation (kindergarten and first grade) and have attended public schools for two years, during second and third grade (*2-year former scholarship recipients*); and (c) 20 students who exited from the program after three years of participation (kindergarten through second grade) and have attended public schools for one year as third graders (*1-year former scholarship recipients*).

Correlational Analysis

A correlational analysis examined whether the amount of time former scholarship recipients have been out of the program attending public schools (ranging from one to three years) is related to achievement or achievement gain. Examining achievement scores at the end of third grade (spring, 2001) revealed that the amount of time former recipients have been out of the program was unrelated to third grade academic achievement. However, achievement gain from late second grade to late third grade on the language and total measures, as well as achievement gain from early first grade to late third grade on the mathematics measure, was positively correlated with the amount of time former scholarship recipients have been out of the program attending public schools. Therefore, by third grade, former recipients who exited the program earlier tended to demonstrate greater gain on these achievement measures in comparison to former recipients who spent more time in the program and exited more recently. Table 41 displays these correlations.

Table 41. Correlations between the Time Former Recipients have been out of the CSTP and Achievement Score Gain

	Correlation Coefficients		
	Language (2nd to 3rd grade)	Mathematics (1st to 3rd grade)	Total (2nd to 3rd grade)
Correlation between (1) achievement gain and (2) duration of non-participation in the CSTP among former recipients	**.227**	**.211**	**.275**
p-value	.025	.037	.006
N	98	98	98

Duration of Non-participation in the CSTP was coded such that larger numbers represent more time out of the program (1 = one year out, 2 = two years out, and 3 = three-years out)

Group Comparisons: Achievement as a Function of Differential Exit

To further examine achievement differences across the four testing episodes among the current and former scholarship recipient groups, mixed-design analyses of variance were conducted on the four achievement measures available from early first grade through the end of third grade. To avoid excluding cases from the already small former scholarship recipient groups, student demographic variables were not included in the analyses as covariates.

Reading Achievement. Descriptive statistics for the analysis of reading achievement change across time are presented in Table 42, and Figure 28 presents these data graphically. Results of the mixed-design analyses of variance on reading achievement are displayed in Table 43. The assumption of sphericity has been violated (Mauchly's W[5] = .820, $p < .001$); therefore, adjustments to the degrees of freedom have been made.

The main effect of testing episode (time) was statistically significant ($p < .001$ using Huynh-Feldt adjustment equal to .897), and follow-up comparisons indicate that, regardless of student group membership (i.e., exit status), achievement in reading increased significantly between successive testing episodes. The main effect of student exit status also was statistically significant. After averaging across the testing episodes, follow-up comparisons found that four-year scholarship recipients, who remained in the program through third grade, achieved at a statistically higher level in reading (M = 581.67) than did either two-year former recipients, who exited after first grade (M = 563.23), or one-year former scholarship recipients, who exited after second grade (M

= 566.54). No other differences in reading achievement among the groups were statistically significant with respect to the main effect of exit status.

Table 42. Student Achievement in Reading from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

Student Group	N	Fall 1st Grade (1998) Mean	SE	Spring 1st Grade (1999) Mean	SE	Spring 2nd Grade (2000) Mean	SE	Spring 3rd Grade (2001) Mean	SE	Marginal Means (Student Group) Mean	SE
4-year Scholarship Recipients	361	**546.19**	1.65	**563.79**	1.99	**598.42**	1.69	**618.26**	2.09	**581.67**	1.46
1-year Former Scholarship Recipients	20	**532.50**	7.02	**543.49**	8.45	**581.57**	7.17	**608.59**	8.86	**566.54**	6.22
2-year Former Scholarship Recipients	40	**528.87**	4.97	**549.50**	5.97	**580.83**	5.07	**593.72**	6.27	**563.23**	4.40
3-year Former Scholarship Recipients	38	**539.88**	5.09	**544.76**	6.13	**593.98**	5.20	**618.62**	6.43	**574.31**	4.51
Marginal Means (Testing Episode)		**536.86**	2.53	**550.38**	3.04	**588.70**	2.58	**609.80**	3.19		

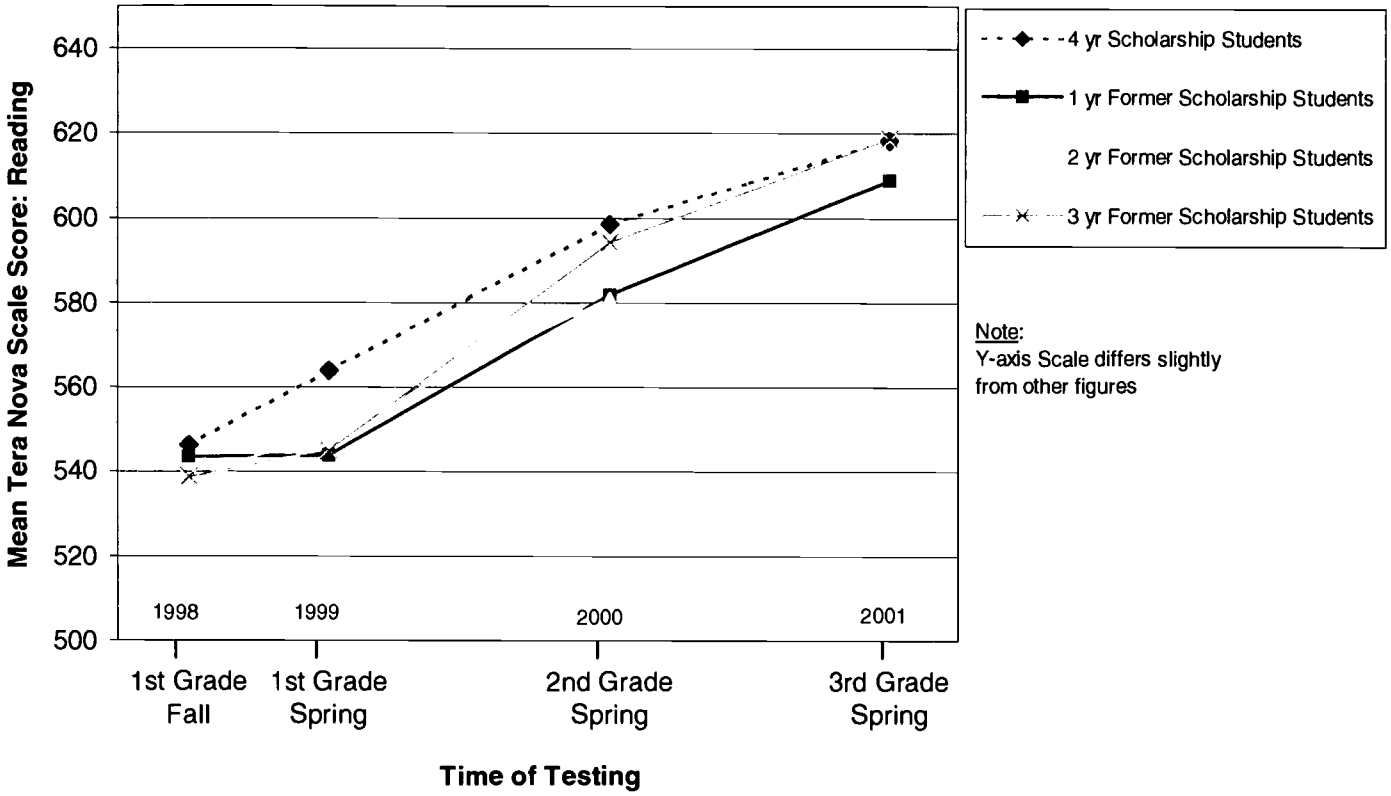
Table 43. Mixed Design Analysis of Variance on Reading Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

Source of Variance	Sum of Squares	Degrees of Freedom[a]	Mean Square	F	p-value	Partial Eta^2 (effect size)[b]
Within Subjects						
Time	523917.82	2.69	194748.34	272.979	< .001	.370
Time x Group	10868.44	8.07	1346.66	1.888	.058	.010
Error	873262.30	1224.06	713.41			
Between Subjects						
Group	65749.14	3	21916.38	7.090	< .001	.040
Error	1406483.23	455	3091.17			

a. Degrees of freedom adjusted due to violation of *sphericity assumption* using Huynh-Feldt adjustment = .897

b. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 28. Reading Achievement of Former Scholarship Recipients vs. Four-year Continuous Scholarship Recipients: 1998 to 2001



In addition, under the assumption of sphericity, the group (exit status) by testing episode interaction was statistically significant. After adjusting the degrees of freedom to account for the violation of this assumption, however, the interaction effect no longer was statistically significant ($p = .058$ using Huynh-Feldt adjustment equal to .897). Therefore, it is most appropriate to conclude that the pattern of achievement gain across the four testing episodes did not differ among the former and current scholarship recipient groups at a statistically significant level.

Language Achievement. Descriptive statistics for the analysis of language achievement change across time are presented in Table 44, and Figure 29 presents these data graphically. Results of the mixed-design analyses of variance on language achievement are displayed in Table 45.[64]

[64] Although, Mauchly's Test of Sphericity ($W[5] = .956$, $p = .001$) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of reading scores still were deemed sufficiently strong (Greenhouse-Geisser = .999 - .970; Huynh-Feldt = .999 - .983). Therefore, no adjustments to degrees of freedom were made.

Table 44. Student Achievement in Language from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the CSTP

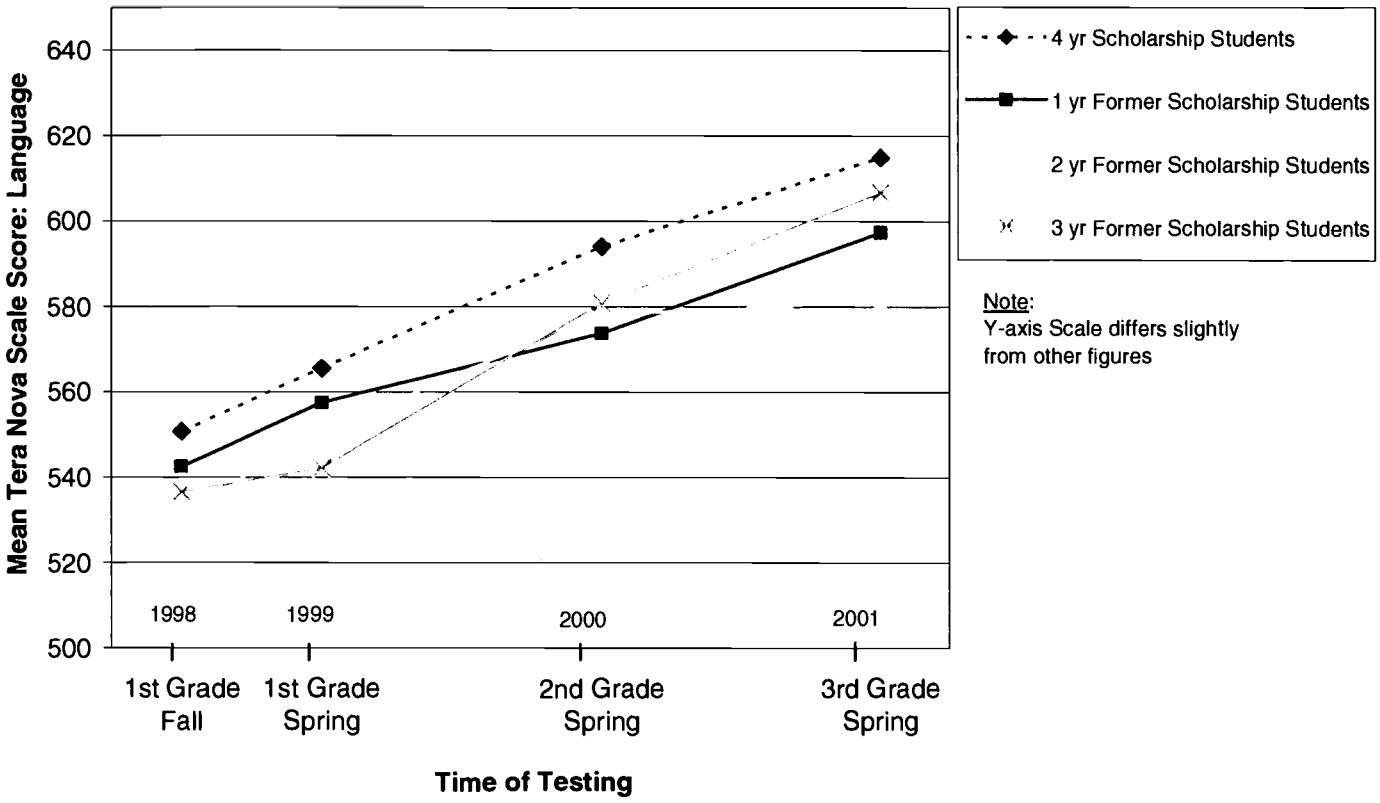
Student Group	N	Fall 1st Grade (1998) Mean	Fall 1st Grade (1998) SE	Spring 1st Grade (1999) Mean	Spring 1st Grade (1999) SE	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student Group) Mean	Marginal Means (Student Group) SE
4-year Scholarship Recipients	361	**550.71**	1.81	**565.43**	1.99	**594.01**	1.90	**614.81**	1.72	**581.24**	1.49
1-year Former Scholarship Recipients	20	**542.19**	7.70	**557.11**	8.44	**573.48**	8.09	**597.02**	7.30	**567.45**	6.31
2-year Former Scholarship Recipients	40	**531.54**	5.45	**550.91**	5.97	**577.85**	5.72	**581.47**	5.16	**560.44**	4.46
3-year Former Scholarship Recipients	38	**536.41**	5.59	**542.03**	6.12	**580.62**	5.87	**606.47**	5.29	**566.38**	4.58
Marginal Means (Testing Episode)		**540.21**	2.77	**553.87**	3.04	**581.49**	2.91	**599.94**	2.63		

Table 45. Mixed Design Analysis of Variance on Language Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	333718.50	3	111239.50	185.496	< .001	.290
Time x Group	15346.55	9	1705.17	2.843	.003	.010
Error	818573.58	1365	599.68			
Between Subjects						
Group	93622.38	3	31207.46	9.798	< .001	.060
Error	1449285.12	455	3185.24			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 29. Language Achievement of Former Scholarship Recipients vs. Four-year Continuous Scholarship Recipients: 1998 to 2001



The main effect of testing episode (time) was statistically significant, and follow-up comparisons indicated that, regardless of student exit status, achievement in language increased significantly between each successive testing episode. The main effect of student exit status also was statistically significant. After averaging language achievement across the four testing episodes, paired comparisons found that four-year scholarship recipients who remained in the program through third grade achieved at a statistically higher level in language (M = 581.24) than did either two-year former recipients who exited the program after first grade (M = 560.44), or three-year former recipients who exited after kindergarten (M = 566.38). No other language achievement differences with respect to the main effect of student exit status were statistically significant.

Unlike reading achievement, the group (exit status) by testing episode interaction was statistically significant, indicating that at least two of the groups displayed different patterns of improvement in language achievement across the testing episodes. In other words, the rate of improvement differed between at least two of the groups. To examine the nature of the interaction, paired comparisons were performed.

At the beginning of first grade, scholarship recipients who remain in the program through third grade achieved at a statistically higher level in language (M = 550.71) than did either recipients who exited after kindergarten (M = 536.41) or recipients who exit after first grade (M = 531.54). Recipients who remain in the program through third grade, however, did not achieve at a statistically different level in language at the beginning of first grade than did students who exit the program after second grade (M = 542.19). By the end of first grade, paired comparisons found that recipients who remain in the program through third grade continued to achieve at a statistically higher level on the language measure (M = 565.43) in comparison to recipients who exited the program after kindergarten (M = 542.03) or recipients who exit the program after first grade (M = 550.91). Notably, recipients who exit at the end of first grade achieved at a lower level in language at the time of their exit in comparison to recipients who remain in the program through third grade.

By the end of second grade, recipients who remain in the program through third grade achieved at a statistically higher level (M = 594.01) than did: recipients who exited after kindergarten (M = 580.62); recipients who exited after first grade (M = 577.85); and recipients who exit the program after second grade (M = 573.48). Notably, the group of recipients who exit the program at the end of second grade achieved at a lower level at the time of their exit in comparison to recipients who remain in the program through third grade.

By the end of third grade, no statistical difference in language achievement was found between recipients who remained in the program (M = 614.81) and recipients who exited after kindergarten (M = 606.47). Recipients who remained in the program at the end of third grade, however, achieved at a statistically higher level on the language measure in comparison to recipients who exited after first grade (M = 581.47) and recipients who exited after second grade

(M = 597.02). Furthermore, by the end of third grade, recipients who exited after kindergarten achieved at a statistically higher level than did recipients who exited after second grade.

Mathematics Achievement. Descriptive statistics for the analysis of mathematics achievement change across time are presented in Table 46, and Figure 30 presents these data graphically. Results of the mixed-design analyses of variance on mathematics achievement are displayed in Table 47.[65]

Table 46. Student Achievement in Mathematics from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

		Fall 1st Grade (1998)		Spring 1st Grade (1999)		Spring 2nd Grade (2000)		Spring 3rd Grade (2001)		Marginal Means (Student group)	
Student group	N	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
4-year Scholarship Recipients	361	**502.10**	1.83	**520.17**	1.67	**552.38**	1.76	**594.27**	1.91	**542.23**	1.46
1-year Former Scholarship Recipients	20	**485.14**	7.75	**502.03**	7.10	**534.75**	7.48	**584.66**	8.12	**526.64**	6.18
2-year Former Scholarship Recipients	40	**486.44**	5.48	**510.09**	5.02	**533.04**	5.29	**569.26**	5.74	**524.71**	4.37
3-year Former Scholarship Recipients	38	**476.68**	5.63	**504.02**	5.15	**552.07**	5.42	**595.95**	5.89	**532.18**	4.49
Marginal Means (Testing Episode)		**487.59**	2.79	**509.08**	2.56	**543.06**	2.69	**586.04**	2.92		

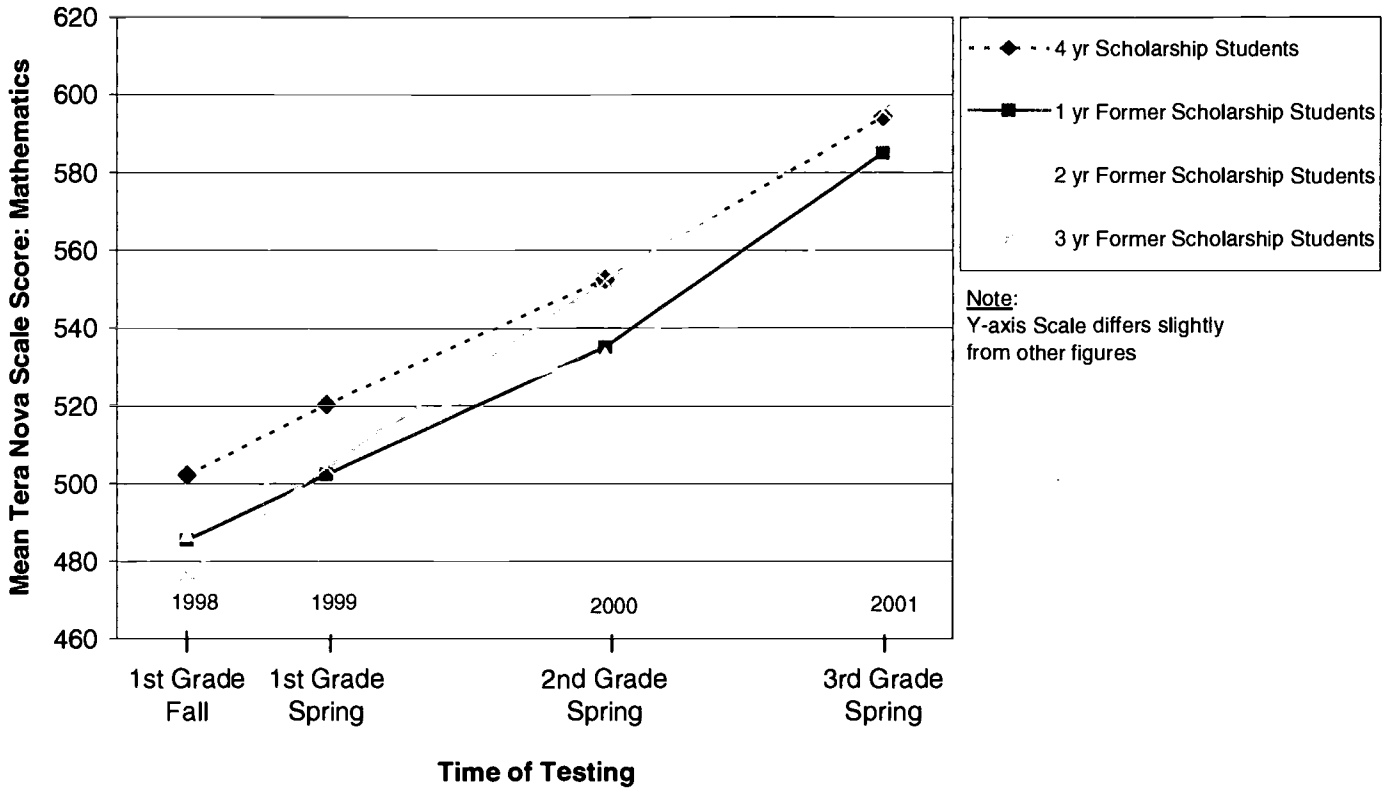
Table 47. Mixed Design Analysis of Variance on Mathematics Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	851391.85	3	283797.28	536.133	< .001	.540
Time x Scholarship Status	23991.24	9	2665.69	5.036	< .001	.030
Error	722550.15	1365	529.34			
Between Subjects						
Scholarship Status	66978.64	3	22326.21	7.301	< .001	.040
Error	1391338.75	455	3057.88			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

[65] Although, Mauchly's Test of Sphericity (W[5] = .956, p = .001) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of reading scores still were deemed sufficiently strong (Greenhouse-Geisser = .999 - .970; Huynh-Feldt = .999 - .983). Therefore, no adjustments to degrees of freedom were made.

Figure 30. Mathematics Achievement of Former Scholarship Recipients vs. Four-year Continuous Scholarship Recipients: 1998 to 2001



The main effect of testing episode (time) was statistically significant, and follow-up comparisons indicated that, regardless of student exit status, achievement in mathematics increased significantly between each successive testing episode. The main effect of student exit status also was statistically significant. Paired follow-up comparisons found that four-year continuing scholarship recipients, on average across the four testing episodes, achieved at a statistically higher level in mathematics (M = 542.23) than did any of the three former scholarship recipient groups: one-year former recipients (M = 526.64); two-year former recipients (M = 524.71), and three-year former scholarship recipients (M = 532.18). No other between-group differences in mathematics achievement were statistically significant with respect to the main effect of exit status.

In addition, the group (exit status) by testing episode interaction was statistically significant, indicating that at least two of the groups displayed different patterns of improvement in

mathematics achievement across the testing episodes. To examine the nature of the interaction, paired comparisons were performed.

Both at the beginning of first grade (M = 502.10) and at the end of first grade (M = 520.17), recipients who remain in the program through third grade achieved at a higher level in mathematics than did: recipients who exited the program after kindergarten (early first grade, M = 476.68; late first grade, M = 504.02); recipients who exit after first grade (early first grade, M = 486.44; late first grade, M = 510.09); and recipients who exit after second grade (early first grade, M = 485.14; late first grade, M = 502.03). As with language achievement, recipients who exit at the end of first grade achieved at a lower level in mathematics at the time of their exit in comparison to recipients who remain in the program through third grade.

By the end of second grade, recipients who remain in the program through third grade continued to achieve at a statistically higher level in mathematics (M = 552.38) than did either recipients who exited after first grade (M = 533.04) or recipients who exit after second grade (M = 534.75). At the end of second grade, however, the difference in mathematics achievement between recipients who remain in the program through third grade and recipients who exited the program after kindergarten (M = 552.07) was no longer statistically significant. Consequently, by the end of second grade, recipients who exited after kindergarten achieved at a statistically higher level than did recipients who exited after first grade and recipients who exit after second grade. As with the language achievement measure, recipients who exit at the end of second grade achieved at a lower level in mathematics at the time of their exit in comparison to recipients who remain in the program through third grade.

By the end of third grade, both recipients who remained in the program (M = 594.27) and recipients who exited after kindergarten (M = 595.95) continued to achieve at a statistically higher level in comparison to recipients who exited after first grade (M = 569.26). However, at the end of third grade, recipients who remained in the program and recipients who exited after kindergarten did not differ significantly in mathematics achievement from each other or the recipients who exited after second grade (M = 584.66).[66]

[66] In addition, it was found that the difference in mathematics achievement between one-year former scholarship recipients and the other former scholarship recipient groups was not statistically significant. Intransitive findings such as this can occur with independent paired comparisons.

Total Achievement. Descriptive statistics for the analysis of total achievement change across time are presented in Table 48, and Figure 31 presents these data graphically. Results of the mixed-design analyses of variance on total achievement are displayed in Table 49. For this analyses, the assumption of sphericity has been violated (Mauchly's W[5] = .796, $p < .001$); therefore, adjustments to the degrees of freedom have been made.

Table 48. Student Total Achievement from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

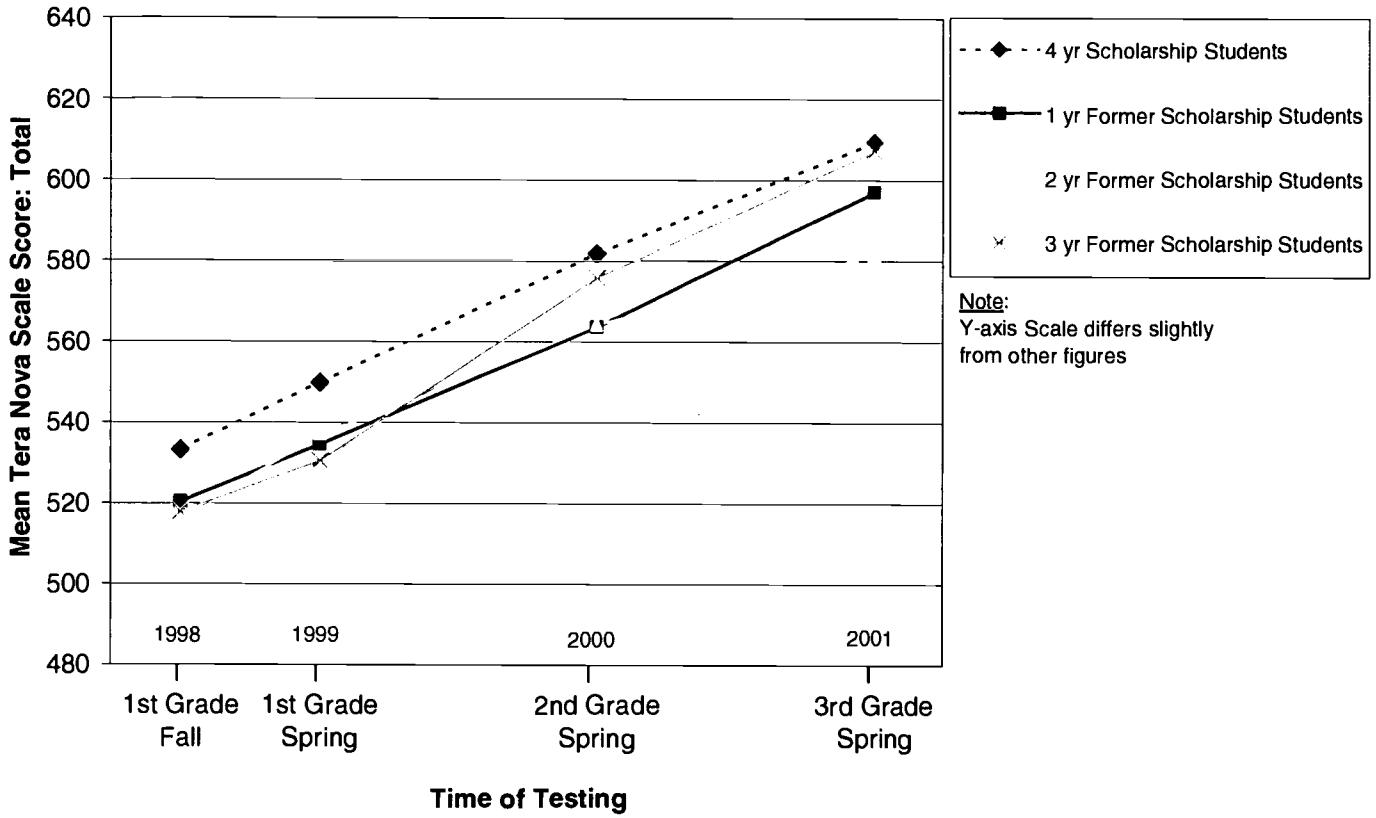
Student group	N	Fall 1st Grade (1998) Mean	Fall 1st Grade (1998) SE	Spring 1st Grade (1999) Mean	Spring 1st Grade (1999) SE	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student group) Mean	Marginal Means (Student group) SE
4-year Scholarship Recipients	361	**533.10**	1.42	**549.73**	1.56	**581.72**	1.54	**609.11**	1.70	**568.41**	1.35
1-year Former Scholarship Recipients	20	**519.94**	6.05	**534.21**	6.61	**563.27**	6.52	**596.75**	7.22	**553.54**	5.75
2-year Former Scholarship Recipients	40	**515.62**	4.28	**536.83**	4.67	**563.90**	4.61	**581.48**	5.11	**549.46**	4.07
3-year Former Scholarship Recipients	38	**517.66**	4.39	**530.27**	4.79	**575.56**	4.73	**607.01**	5.24	**557.62**	4.17
Marginal Means (Testing Episode)		**521.58**	2.18	**537.76**	2.38	**571.11**	2.35	**598.59**	2.60		

Table 49. Mixed Design Analysis of Variance on Total Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001) based on Differential Exit from the Program

Source of Variance	Sum of Squares	Degrees of Freedom[a]	Mean Square	F	p-value	Partial Eta^2 (effect size)[b]
Within Subjects						
Time	546240.52	2.61	209220.74	637.452	< .001	.580
Time x Scholarship Status	12340.17	7.83	1575.51	4.800	< .001	.030
Error	389894.86	1187.93	328.21			
Between Subjects						
Scholarship Status	74166.51	3	24722.17	9.352	< .001	.050
Error	1202812.11	455	2643.54			

a. Degrees of freedom adjusted due to violation of *sphericity assumption* using Huynh-Feldt adjustment = .870
b. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 31. Total Achievement of Former Scholarship Recipients vs. Four-year Continuous Scholarship Recipients: 1998 to 2001



The main effect of testing episode (time) was statistically significant ($p < .001$ using Huynh-Feldt adjustment equal to .87) indicating that, regardless of student exit status, total achievement increased at a statistically significant level between each successive testing episode. The main effect of student exit status was also statistically significant. Paired comparisons found that four-year continuing scholarship recipients, on average across the four testing episodes, achieved at a statistically higher level ($M = 568.41$) than did any of the three former scholarship recipient groups: one-year former recipients ($M = 553.54$); two-year former recipients ($M = 549.46$), and three-year former scholarship recipients ($M = 557.62$). No other between-group differences in total achievement were statistically significant with respect to the main effect of student exit status.

In addition, the group (exit status) by testing episode interaction was statistically significant after accounting for the violation of the sphericity assumption ($p < .001$ using the Huynh-Feldt adjustment equal to .87). Paired comparisons were again conducted.

Both at the beginning (M = 533.10) and at the end of first grade (M = 549.73), follow-up comparisons revealed that recipients who remain in the program through third grade achieved at a higher level than did: recipients who exited after kindergarten (early first grade, M = 517.66; late first grade, M = 530.27); recipients who exit after first grade (early first grade, M = 515.62; late first grade, M = 536.83); and recipients who exit after second grade (early first grade, M = 519.94; late first grade, M = 534.21). Notably, recipients who exit at the end of first grade achieved at a lower level at the time of their exit in comparison to recipients who remain in the program through third grade.[67]

By the end of second grade, recipients who remain in the program through third grade continued to achieve at a statistically higher level on the total achievement measure (M = 581.72) in comparison to recipients who exited from the program after first grade (M = 563.90) and recipients who exit the program after second grade (M = 563.27).[32] However, the difference in total achievement at the end of second grade between recipients who remain in the program through third grade (M = 581.72) and recipients who exited after kindergarten (M = 575.56) was no longer statistically significant. Furthermore, at the end of second grade, recipients who exited after kindergarten achieved at a statistically higher level than did either recipients who exited after first grade or recipients who exit after second grade. Notably, recipients who exit at the end of second grade achieved at a lower level at the time of their exit in comparison to recipients who remain in the program through third grade.

By the end if third grade, recipients who remained in the program continued to achieve at a statistically higher level on the total achievement measure (M = 609.11) than did either recipients who exited after first grade (M = 581.48) or recipients who exited after second grade (M = 596.75). However, the difference in total achievement between recipients who remained in the program at the end of third grade and recipients who exited after kindergarten (M = 607.01) was not statistically significant, and recipients who exited after second grade (M = 596.75) were no longer statistically different from recipients who exited after kindergarten (M = 607.01). Furthermore, at the end of third grade, recipients who exited after kindergarten and after second

[67] All p-values < .001 using Huynh-Feldt adjustment in degrees of freedom equal to .870.

grade were achieving at a statistically higher level in comparison to the recipients who exited after first grade (M = 581.48).

Achievement Gain from First Grade to Third Grade: A Comparison of Three-year Former Scholarship Recipients and Four-year Continuous Scholarship Recipients

Comparisons of achievement gain over three academic years were conducted examining the change in language, mathematics, and total achievement scores from the end of first grade to the end of third grade between (a) three-year former recipients, who exited the program after kindergarten to attend public schools through third grade and (b) four-year continuous scholarship recipients, who remained in the program from kindergarten through third grade. Reading scores were not included in this analysis because the student group by testing episode interaction was not statistically significant on the reading measure.

From the end of first grade to the end of third grade, former recipients who exited the CSTP after kindergarten displayed a significantly greater gain on all of the achievement measures analyzed in comparison to scholarship recipients who remained in the program continuously through third grade. That is, among the scholarship recipients who entered the program in kindergarten (1997-1998), students who exited to attend public schools for the next three years gained significantly more in language, mathematics, and total achievement than did students who remained in the program. This differential gain in academic achievement eliminated the significant differences between continuing recipients and former recipients that were found at the end of first grade (see the analyses reported above). Consequently, by the end of third grade, continuing recipients and former recipients who exited after kindergarten achieved at similar levels on the language, mathematics, and total measures. Table 50 below displays the mean scale scores and gain scores for these groups from the end of first grade to the end of third grade.

Orthogonal Comparisons of the Differential Exit Groups and Continuous Scholarship Recipients

Additional follow-up comparisons were conducted to simplify the examination of achievement differences between continuing scholarship recipients and the former recipient groups based on differential exit from the CSTP. Specifically, orthogonal comparisons were performed within and between each of the four testing episodes comparing students who, at each episode, remained in the program with students who, during that same episode, had exited the program

and were attending public schools. In other words, for each testing episode, only two groups were compared: (a) students who were former scholarship recipients attending public schools at the time and (b) students who were active scholarship users attending private schools at the time.[68] For example, *at the end of second grade*, former scholarship recipients include students who exited the program after kindergarten and first grade, and continuous recipients include students who will remain in the program through third grade as well as recipients who will exit the program after second grade. To avoid excluding cases from the already small former scholarship recipient groups, student demographic variables were not included in the analyses as covariates.

Table 50. Academic Achievement from First Grade to Third Grade: Three-year Former Scholarship Recipients versus Four-year Continuous Scholarship Recipients

Achievement Measure	Student Exit Status	Scale Score (Mean)		Gain Score (Mean)[c]
		End of 1st grade[a]	End of 3rd grade[b]	
Language	4-year Scholarship Recipients	565.43	614.81	49.38
	3-year Former Recipients	542.03	606.47	64.44
Mathematics	4-year Scholarship Recipients	520.17	594.27	74.10
	3-year Former Recipients	504.02	595.95	91.93
Total Achievement	4-year Scholarship Recipients	549.73	609.11	59.38
	3-year Former Recipients	530.27	607.01	76.74

4-year Scholarship Recipients, N = 361. 3-year Former Scholarship Recipients, N = 38

a. All mean differences at the end of 1st grade are statistically significant.
b. All mean differences at the end of 3rd grade are not statistically significant.
c. 3-year Former recipients' mean gain is significantly greater than 4-year Scholarship recipients' mean gain on all of the achievement measures in the table.

Language Achievement. During first grade, three of the comparison groups remained in the program (recipients who will remain in the program through third grade, recipients who will exit after first grade, and recipients who will exit after second grade) and one group consisted of former recipients who had left the program after kindergarten. At the beginning of first grade, no significant difference was found between students in the program (M = 541.48) and those who had left the program and were attending public schools (M = 536.41). Student who had left the program after kindergarten were achieving at similar levels as those who remained. From the

[68] The former scholarship recipient and current/continuing scholarship recipient groups were created by averaging across multiple sub-groups of students who comprised these groups during a given testing episode.

beginning to the end of first grade, the language achievement gain displayed by former scholarship recipients who exited after kindergarten (MΔ = 5.62) did not differ significantly from the gain displayed by recipients who remained in the program (MΔ = 16.34). By the end of first grade, however, students who continued to use a scholarship to attend private schools had gained enough to achieve at a significantly higher level on the language measure (M = 557.82) in comparison to former recipients attending public schools at the time (M = 542.03).

From the end of first grade to the end of second grade, former scholarship recipients (students who exited the program after kindergarten and after first grade) displayed a significantly greater gain in language achievement (MΔ = 32.77) than did recipients who remained in the program (MΔ = 22.48). Consequently, by the end of second grade, the difference between students who remained in the program (four-year continuous recipients and recipients who will exit after second grade) and those who, at this point, had exited the program and were attending public schools was not statistically significant (M = 583.75 and M = 579.23, respectively). From the end of second grade to the end of third grade, the gain in language achievement displayed by former scholarship recipients who exited after kindergarten, first grade, and second grade (MΔ = 17.67) did not differ significantly from the gain displayed by recipients who remained in the program through third grade (MΔ = 20.80). However, despite a statistically similar gain, language achievement for the remaining scholarship students (M = 614.81) was significantly higher at the end of third grade in comparison to the three groups of former recipients who had exited the program in previous years combined (M = 594.99).

Mathematics Achievement. At the beginning of first grade, the mathematics achievement of recipients who remained in the program was significantly higher than the achievement of recipients who exited the program after kindergarten and were attending public schools (M = 491.23 and 476.68, respectively). From the beginning to the end of first grade, both continuing recipients and former recipients made statistically similar gains in mathematics achievement (MΔ = 19.53 and 27.34, respectively). As a result, the higher level of achievement at the beginning of first grade remained significant for continuing recipients in comparison to former students at the end of first grade (M = 510.76 and 504.02).

From the end of first grade to the end of second grade, recipients who remained in the program (recipients who participate through third grade and recipients who will exit after second grade) did not gain significantly more on the mathematics measure (MΔ = 32.47) than did former recipients at the time (MΔ = 35.50). By the end of second grade, there was no statistically significant difference in mean mathematics achievement between the recipients who remained in the program (M = 543.57) and the recipients who exited the program to attend public schools (M = 542.56). From the end of second grade to the end of third grade, recipients who remained in the program did not gain significantly more on the mathematics measure (MΔ = 41.89) in comparison to the three groups of former participants combined (MΔ = 43.34). As in the area of language achievement, recipients who remained in the program through third grade (M = 594.27) were achieving at significantly higher levels in mathematics in comparison to all of the recipients who exited the program in previous years (M = 583.29).

Total Achievement. At the beginning of first grade, students remaining in the program had similar total achievement (M = 522.89) to students who exited the program after kindergarten and enrolled in public schools (M = 517.66). From the beginning to the end of first grade, both continuing recipients and former recipients made statistically similar gains in total achievement (MΔ = 17.37 and 12.61, respectively). By the end of first grade, however, the total achievement of continuing recipients was significantly higher (M = 540.26) than that of former recipients (M = 530.27).

From the end of first grade to the end of second grade, recipients who remained in the program (recipients who participate through third grade and recipients who will exit after second grade) did not gain significantly more on the mathematics measure (MΔ = 30.53) than did former recipients at the time (MΔ = 36.18). At the end of second grade, the difference in total achievement between recipients remaining in the program and those who exited to attend public schools was no longer significant (M = 572.50 and 569.73, respectively). From the end of second grade to the end of third grade, recipients who remained in the program did not gain significantly more on the mathematics measure (MΔ = 27.39) in comparison to the three groups of former participants combined (MΔ = 27.50). However, recipients who remained in the program through third grade had total achievement scores (M = 609.11) that were significantly

higher than those of the three groups of recipients who had exited the program in previous years combined (M = 595.08).

Achievement of Differential Exit Groups versus Public School Comparison Groups: Orthogonal Comparisons

Mixed-model analyses of variance were conducted to compare the academic achievement of the former scholarship recipient groups (one-year, two-year, and three-year former recipients) to the public school groups (applicant non-recipients, non-applicants, and scholarship winner non-users).[69] The purpose of these analyses is to examine whether students who exited the CSTP at various points from kindergarten to third grade were similar to or different from any of the public school comparison groups in terms of academic achievement. Separate analyses were performed for each of the four measures of academic achievement. Because comparisons of achievement among the former scholarship recipient groups, as well as among the public school comparison groups, already have been conducted and discussed above, the following discussion focuses on the similarities and differences *between* the former scholarship recipient groups and the public school groups. For clarification, however, some key differences that have been found and discussed previously are briefly highlighted again.

Overall, the analyses of the four measures of academic achievement revealed few differences in academic achievement among the public school comparison groups, either within or between testing episodes. Students in the public school comparison groups achieved at statistically similar levels within each testing episode, and they displayed similar amounts of improvement in achievement across time. Consequently, the achievement differences between former scholarship recipients and the public school comparison groups were similar for applicant non-recipients, non-applicants, and scholarship winner non-users. Therefore, the separate public school groups have been aggregated to simplify the follow-up analyses. Together, these groups are referred to as continuous or continuing public school students in the analyses presented below.

Furthermore, the analyses revealed that the main effect of testing episode (time) was statistically significant for each measure – indicating that achievement increased significantly from one

[69] To avoid excluding cases from the already small former scholarship recipient groups in an effort to maximize the statistical power of the analyses, student demographic variables (sex, minority status, and estimated meal code) were not included in the analyses as covariates.

testing episode to the next, regardless of group membership. Therefore, the main effect of testing episode is not discussed below for each achievement measure. In addition, because the group by testing episode interaction qualifies the presence or absence of a statistically significant main effect of group membership, and because this interaction was statistically significant for each of the four measures, the main effect of group membership is not discussed below. Instead, orthogonal pairwise comparisons were conducted at each testing episode contrasting the achievement of: (a) students who were former scholarship recipients attending public schools at the time, (b) students who eventually exit the program but were active scholarship users attending private schools at the time, and (c) continuing public school students.

Reading Achievement. Descriptive statistics for the analysis of reading achievement across time are presented in Table 51, and Figure 32 presents these data graphically. Results of the mixed-design analysis of variance on reading achievement are displayed in Table 52.[70]

Table 51. Student Achievement in Reading from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

		Fall 1st Grade (1998)		Spring 1st Grade (1999)		Spring 2nd Grade (2000)		Spring 3rd Grade (2001)		Marginal Means (Student group)	
Student group	N	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1-year Former Scholarship Recipients	20	**532.50**	7.69	**543.49**	8.42	**581.57**	7.72	**608.59**	8.65	**566.54**	6.39
2-year Former Scholarship Recipients	40	**528.87**	5.44	**549.50**	5.95	**580.83**	5.46	**593.72**	6.12	**563.23**	4.52
3-year Former Scholarship Recipients	38	**539.88**	5.58	**544.76**	6.11	**593.98**	5.60	**618.62**	6.28	**574.31**	4.64
Public Applicant Non-recipient	419	**536.13**	1.68	**562.91**	1.84	**595.78**	1.69	**613.04**	1.89	**576.97**	1.40
Public Scholarship Winner Non-user	85	**535.63**	3.73	**566.16**	4.08	**594.43**	3.74	**613.09**	4.20	**577.33**	3.10
Public Non-applicant	513	**531.62**	1.52	**562.18**	1.66	**594.51**	1.52	**612.75**	1.71	**575.26**	1.26
Marginal Means (Testing Episode)		**534.10**	1.96	**554.83**	2.14	**590.18**	1.97	**609.97**	2.21		

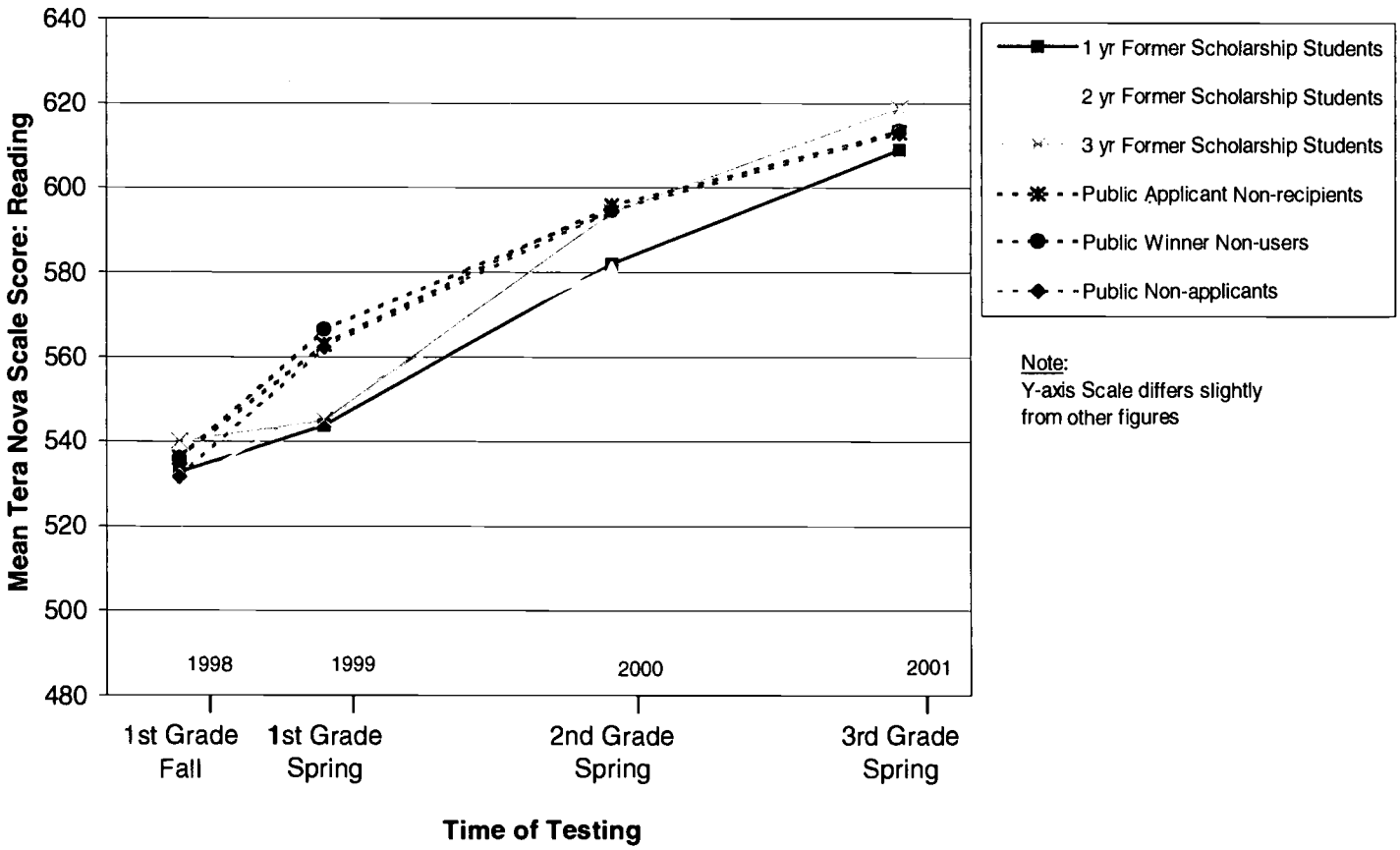
[70] Although, Mauchly's Test of Sphericity (W[5] = .861, $p < .001$) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of reading scores still were deemed sufficiently strong (Greenhouse-Geisser = .999 - .9220; Huynh-Feldt = .999 - .928). Therefore, no adjustments to degrees of freedom were made.

Table 52. Mixed Design Analysis of Variance on Reading Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	1073932.29	3	357977.43	532.34	< .001	.320
Time x Scholarship Status	25084.28	15	1672.29	2.49	.001	.010
Error	2237269.20	3327	672.45			
Between Subjects						
Scholarship Status	35567.24	5	7113.45	2.176	.055	.010
Error	3625399.66	1109	3269.07			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 32. Reading Achievement of Former Scholarship Recipients vs. Public School Comparison Groups: 1998 to 2001



At the beginning of first grade, former recipients who exited the CSTP after kindergarten to attend public schools achieved at a statistically similar level on the reading measure (M = 539.88) in comparison both to continuing scholarship recipients who remained in the program, but would exit in future years, (M = 530.69) and to their continuing public school classmates (M = 534.46). Furthermore, there was no difference in reading achievement between continuing scholarship recipients and continuing public school students at the beginning of first grade. At the end of first grade, however, continuing public school students achieved at a significantly higher level (M = 563.75) than both former scholarship recipients (M = 544.76) and continuing scholarship recipients who would exit in future years (M = 546.5). Former recipients and recipients who remained in the program achieved at a statistically similar level on the reading measure at the end of first grade.

At the end of second grade, continuing public school students sustained a significantly higher level of reading achievement (M = 594.91) in comparison both to continuing scholarship recipients who would exit after second grade (M = 581.57) and to former scholarship recipients who exited the program after kindergarten and first grade (M = 587.45). Former recipients and recipients who remained in the program, again, achieved at a statistically similar level on the reading measure at the end of second grade. Similarly, by the end of third grade, continuing public school students (M = 612.96) achieved at a significantly higher level in comparison to all former recipients who had exited the CSTP to attend public schools (M = 606.98).

Language Achievement. Descriptive statistics for the analysis of language achievement across time are presented in Table 53, and Figure 33 presents these data graphically. Results of the mixed-design analyses of variance on language achievement are displayed in Table 54.[71]

Orthogonal comparisons revealed no significant differences between the student groups from the beginning of first grade through the end of second grade. However, by the end of third grade, students who continuously attended public schools through third grade (applicant non-recipients, non-applicants, and winner non-users) had significantly higher language achievement (M =

[71] Although, Mauchly's Test of Sphericity (W[5] = .938, $p < .001$) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of language scores still were deemed sufficiently strong (Greenhouse-Geisser = .999 - .962; Huynh-Feldt = .999 - .969). Therefore, no adjustments to degrees of freedom were made.

606.96) in comparison to students who exited the CSTP after one or more of the preceding years (M = 594.99).

Table 53. Student Achievement in Language from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

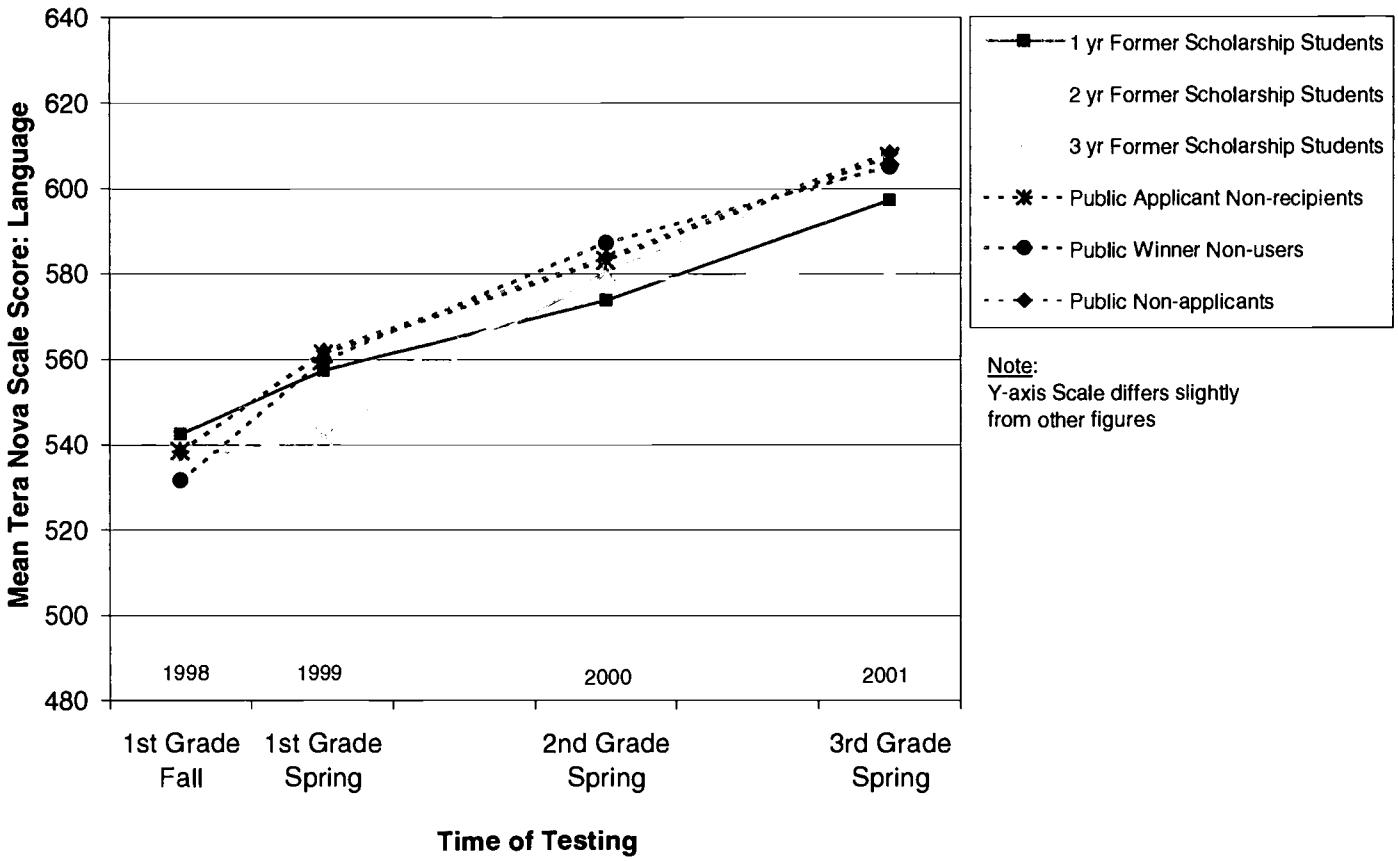
Student group	N	Fall 1st Grade (1998) Mean	Fall 1st Grade (1998) SE	Spring 1st Grade (1999) Mean	Spring 1st Grade (1999) SE	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student group) Mean	Marginal Means (Student group) SE
1-year Former Scholarship Recipients	20	**542.19**	7.96	**557.11**	8.63	**573.48**	7.84	**597.02**	7.63	**567.45**	6.41
2-year Former Scholarship Recipients	40	**531.54**	5.63	**550.91**	6.11	**577.85**	5.55	**581.47**	5.40	**560.44**	4.53
3-year Former Scholarship Recipients	38	**536.41**	5.78	**542.03**	6.26	**580.62**	5.69	**606.47**	5.54	**566.38**	4.65
Public Applicant Non-recipient	419	**538.49**	1.74	**561.27**	1.89	**583.12**	1.71	**607.50**	1.67	**572.59**	1.40
Public Scholarship Winner Non-user	85	**531.33**	3.86	**559.47**	4.19	**586.97**	3.80	**604.95**	3.70	**570.68**	3.12
Public Non-applicant	513	**538.18**	1.57	**561.91**	1.71	**583.69**	1.55	**608.42**	1.51	**573.05**	1.27
Marginal Means (Testing Episode)		**536.35**	2.03	**555.45**	2.20	**580.95**	2.00	**600.97**	1.94		

Table 54. Mixed Design Analysis of Variance on Language Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	739846.87	3	246615.62	396.504	< .001	.260
Time x Scholarship Status	26724.53	15	1781.64	2.864	< .001	.010
Error	2069312.83	3327	621.97			
Between Subjects						
Scholarship Status	31244.60	5	6248.92	1.901	.091	< .001
Error	3645396.32	1109	3287.10			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 33. Language Achievement of Former Scholarship Recipients vs. Public School Comparison Groups: 1998 to 2001



Mathematics Achievement. Descriptive statistics for the analysis of mathematics achievement change across time are presented in Table 55, and Figure 34 presents these data graphically. Results of the mixed-design analyses of variance on mathematics achievement are displayed in Table 56.[72]

[72] Although, Mauchly's Test of Sphericity (W[5] = .972, p < .001) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of mathematics scores still were deemed sufficiently strong (Greenhouse-Geisser = .999 - .981; Huynh-Feldt = .999 - .988). Therefore, no adjustments to degrees of freedom were made.

Table 55. Student Achievement in Mathematics from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

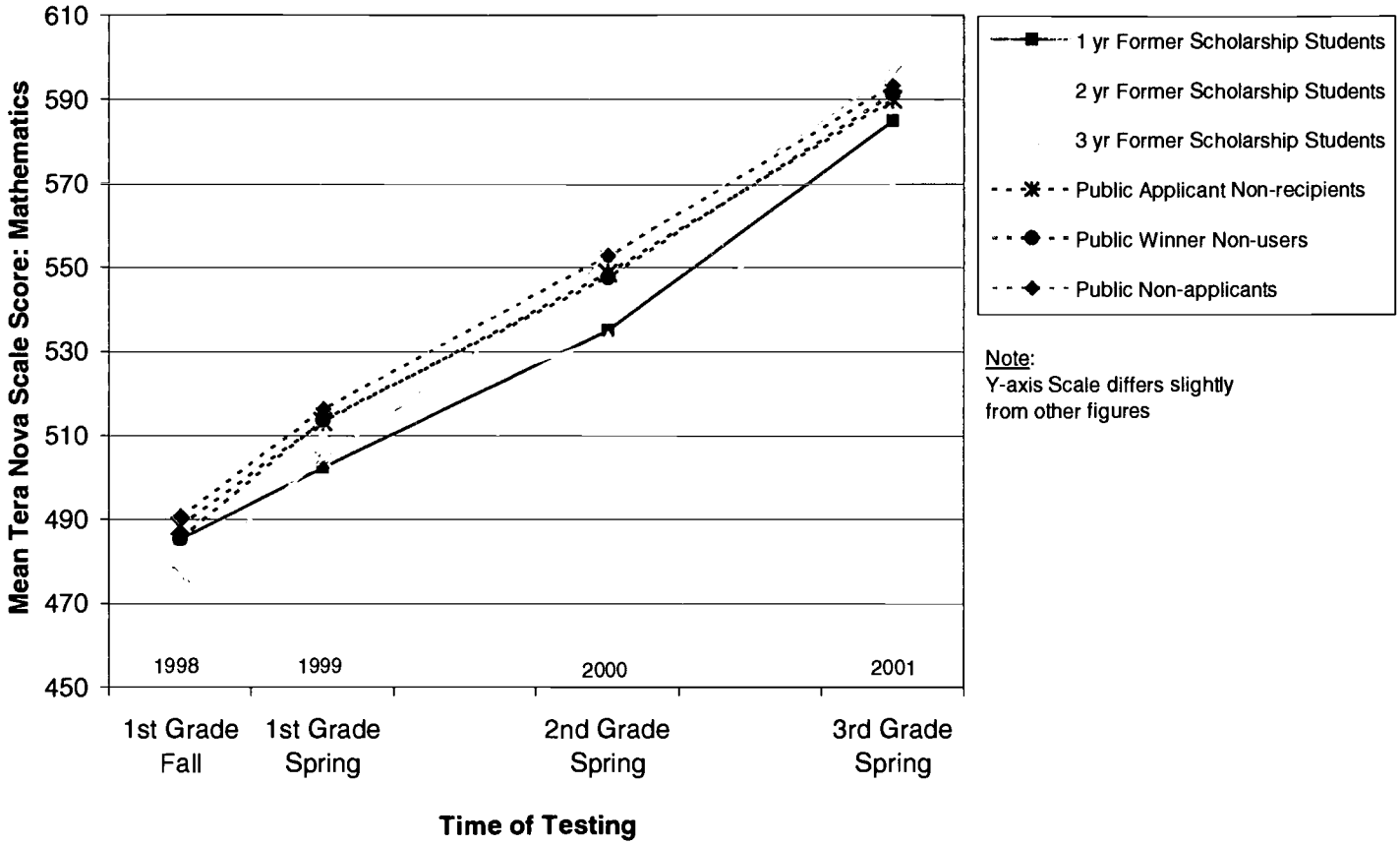
Student group	N	Fall 1st Grade (1998) Mean	Fall 1st Grade (1998) SE	Spring 1st Grade (1999) Mean	Spring 1st Grade (1999) SE	Spring 2nd Grade (2000) Mean	Spring 2nd Grade (2000) SE	Spring 3rd Grade (2001) Mean	Spring 3rd Grade (2001) SE	Marginal Means (Student group) Mean	Marginal Means (Student group) SE
1-year Former Scholarship Recipients	20	**485.14**	7.48	**502.03**	8.16	**534.75**	8.24	**584.66**	8.76	**526.64**	6.77
2-year Former Scholarship Recipients	40	**486.44**	5.29	**510.09**	5.77	**533.04**	5.82	**569.26**	6.19	**524.71**	4.79
3-year Former Scholarship Recipients	38	**476.68**	5.42	**504.02**	5.92	**552.07**	5.98	**595.95**	6.36	**532.18**	4.91
Public Applicant Non-recipient	419	**488.47**	1.63	**513.32**	1.78	**548.66**	1.80	**590.01**	1.91	**535.11**	1.48
Public Scholarship Winner Non-user	85	**485.21**	3.63	**513.48**	3.96	**547.42**	4.00	**591.05**	4.25	**534.29**	3.29
Public Non-applicant	513	**490.67**	1.48	**516.32**	1.61	**552.70**	1.63	**593.31**	1.73	**538.25**	1.34
Marginal Means (Testing Episode)		**485.44**	1.91	**509.87**	2.08	**544.77**	2.10	**587.37**	2.23		

Table 56. Mixed Design Analysis of Variance on Mathematics Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	1804965.01	3	601655.00	1080.614	< .001	.490
Time x Scholarship Status	22077.71	15	1471.85	2.644	.001	.010
Error	1852379.09	3327	556.77			
Between Subjects						
Scholarship Status	42250.23	5	8450.05	2.302	.043	.010
Error	4070754.97	1109	3670.65			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 34. Mathematics Achievement of Former Scholarship Recipients vs. Public School Comparison Groups: 1998 to 2001



Students who continuously attended public schools from kindergarten through third grade performed at significantly higher levels in mathematics at each testing episode in comparison to former recipients who exited the CSTP to attend public schools at various points during this period. Specifically, at the beginning of first grade, as well as at the end of first, second, and third grade, students who continuously attended public schools achieved at significantly higher levels on the mathematics measure (M = 488.12, 514.37, 549.59, and 591.46, respectively) in comparison to former scholarship recipients who exited the program in preceding years to attend public schools (M = 476.68, 504.02, 542.56, and 583.29, respectively). Furthermore, at the end of both first and second grade, students who continuously attended public schools also achieved at significantly higher levels on the mathematics measure (M = 514.37 and 549.59, respectively) in comparison to continuing scholarship recipients who would eventually exit the program to attend public schools (M = 506.06 and 534.75, respectively). Only at the beginning of first grade

did students who continuously attended public schools (M = 488.12) achieve at a statistically similar level to continuing scholarship recipients who would eventually exit the program (M = 485.79). Furthermore, from the beginning of first grads to the end of second grade, former scholarship recipients and continuing scholarship recipients, who would eventually exit the program, did not differ significantly in mathematics achievement.

Total Achievement. Descriptive statistics for the analysis of total achievement change across time are presented in Table 57, and Figure 35 presents these data graphically. Results of the mixed-design analyses of variance on total achievement are displayed in Table 58.[73]

Table 57. Student Total Achievement in from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

		Fall 1st Grade (1998)		Spring 1st Grade (1999)		Spring 2nd Grade (2000)		Spring 3rd Grade (2001)		Marginal Means (Student group)	
Student group	N	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1-year Former Scholarship Recipients	20	**519.94**	6.40	**534.21**	7.14	**563.27**	6.91	**596.75**	7.43	**553.54**	6.06
2-year Former Scholarship Recipients	40	**515.62**	4.53	**536.83**	5.05	**563.90**	4.89	**581.48**	5.25	**549.46**	4.28
3-year Former Scholarship Recipients	38	**517.66**	4.65	**530.27**	5.18	**575.56**	5.01	**607.01**	5.39	**557.62**	4.39
Public Applicant Non-recipient	419	**521.01**	1.40	**545.83**	1.56	**575.67**	1.51	**603.52**	1.62	**561.51**	1.32
Public Scholarship Winner Non-user	85	**517.39**	3.11	**546.37**	3.46	**576.27**	3.35	**603.03**	3.60	**560.77**	2.94
Public Non-applicant	513	**520.14**	1.26	**546.81**	1.41	**577.08**	1.37	**604.81**	1.47	**562.21**	1.20
Marginal Means (Testing Episode)		**518.63**	1.63	**540.05**	1.82	**571.96**	1.76	**599.43**	1.89		

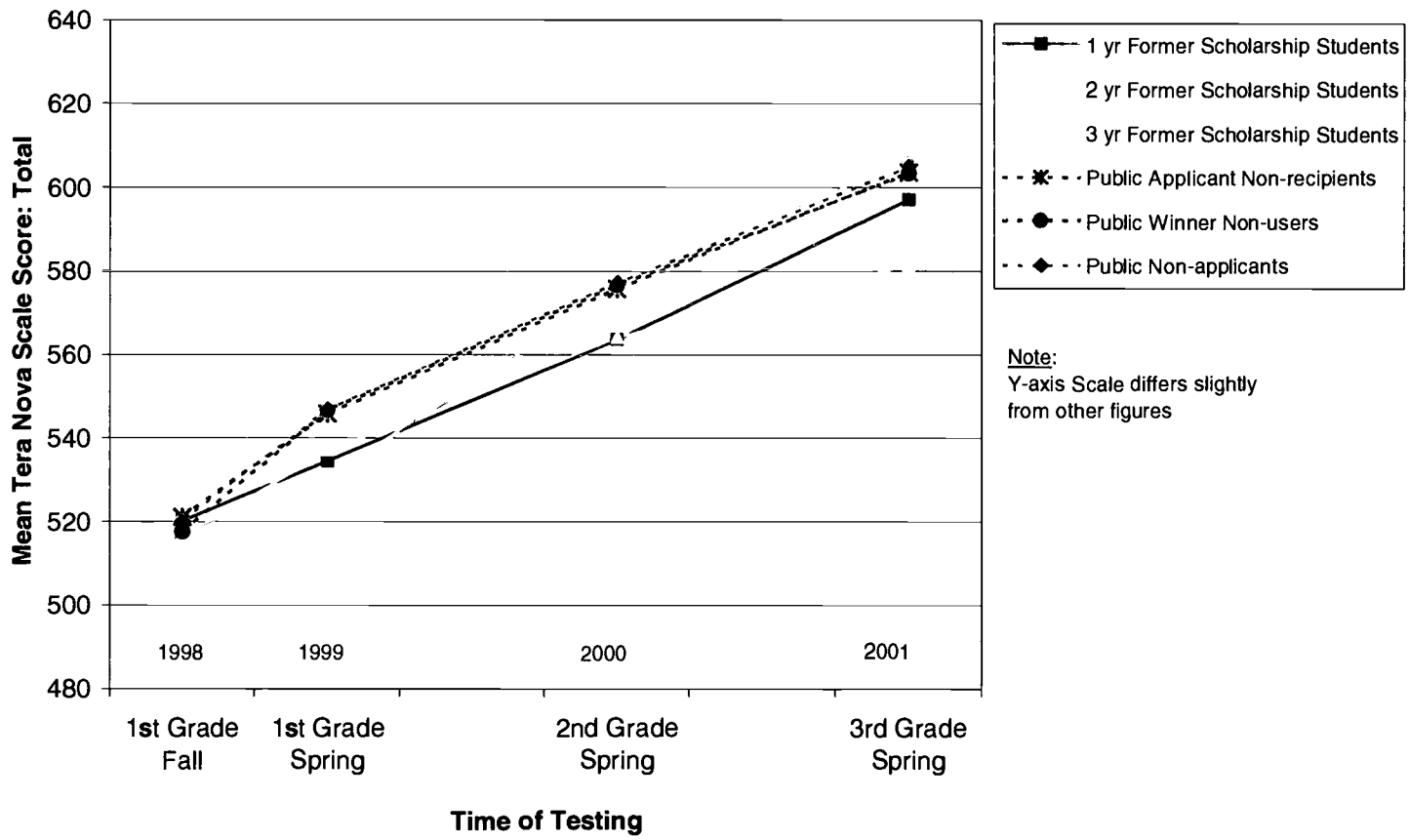
[73] Although, Mauchly's Test of Sphericity (W[5] = .914, p < .001) revealed a violation of the sphericity assumption, estimates of sphericity for the analysis of total achievement scores still were deemed sufficiently strong (Greenhouse-Geisser = .999 - .947; Huynh-Feldt = .999 - .954). Therefore, no adjustments to degrees of freedom were made.

Table 58. Mixed Design Analysis of Variance on Total Achievement Scale Scores from Early First Grade (Fall 1998) to Late Third Grade (Spring 2001): Differential Exit from the CSTP vs. Public School Comparison Groups

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value	Partial Eta^2 (effect size)[a]
Within Subjects						
Time	1159877.25	3	386625.75	1201.933	< .001	.520
Time x Scholarship Status	17644.29	15	1176.29	3.657	< .001	.010
Error	1070196.16	3327	321.67			
Between Subjects						
Scholarship Status	30930.44	5	6186.09	2.109	.062	< .001
Error	3253265.93	1109	2933.51			

a. Partial eta squared is the ratio of the variation accounted for by an individual independent variable (SS:IV) to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole (SS:IV+SSE). Interpretation: Percent of variance in achievement scores accounted for by corresponding effect.

Figure 35. Total Achievement of Former Scholarship Recipients vs. Public School Comparison Groups: 1998 to 2001



No significant differences in total achievement were found among the groups at the beginning of first grade. By the end of first grade, however, students who continuously attended public schools since kindergarten had higher total achievement (M = 546.34) in comparison both to former scholarship recipients who exited the program after kindergarten (M = 530.27) and to scholarship recipients who remained in the program, but would leave in subsequent years (M = 535.52). Similarly, at the end of second grade, continuous public school students had higher total achievement (M = 576.34) in comparison both to former scholarship students who exited the program after kindergarten and first grade (M = 569.73) and to continuing scholarship recipients who eventually exit the program after second grade (M = 563.27). However, at the end of both first and second grade, former scholarship recipients and continuing scholarship recipients, who eventually exit the program, did not differ significantly on the total achievement measure. Finally, by the end of third grade, students who continuously attended public schools since kindergarten had significantly higher total achievement (M = 603.79) in comparison to all former recipients who exited the CSTP to attend public schools over the preceding years combined (M = 595.08).

Achievement Gain from First Grade to Third Grade: Comparison of Three-year Former Scholarship Recipients, Public School Students, and Students who Exit the CSTP after First or Second Grade

At the end of first grade, both (a) former scholarship recipients who exited from the program after kindergarten and (b) students in the CSTP at the time, who eventually would exit the program after first or second grade, achieved at a significantly lower level on all of the achievement measures in comparison to their peers in public schools. However, current and future former scholarship recipients did not differ on any of the achievement measures at the end of first grade.

From the end of first grade to the end of third grade, former scholarship recipients – who exited after kindergarten to attend public schools – gained more on all of the achievement measures than did either (a) their classmates in public schools or (b) former scholarship recipients who remained in the program longer and exited after first or second grade. Continuous public school students and former recipients who exited after first or second grade did not differ in the amount of achievement score gain they displayed from first to third grade on the reading, mathematics,

and total measures. On the language measure, continuous public school students displayed a significantly greater achievement score gain than did former recipients who exited after first or second grade.

Consequently, by the end of third grade, former recipients – who exited the program after kindergarten to attended public schools through third grade – and continuous public school students achieved at a statistically similar level on all of the measures. That is, the greater gain in achievement displayed by three-year former scholarship recipients in comparison to their classmates in public schools eliminated the significant difference in achievement scores that existed between these groups at the end of first grade. The other former scholarship recipients – who remained in the program longer and exited after either first or second grade – continued to achieve at a lower level on all of the measures in comparison to continuous public school students at the end of third grade. Table 59 displays the mean achievement scores for these groups across all of the measures.

Table 59. Academic Achievement: Three-year Former Scholarship Recipients versus Public School Students and Other Former Scholarship Recipients

Achievement Measure	Student Exit Status	Scale Score[a] (Mean) End of 1st grade	Scale Score[a] (Mean) End of 3rd grade	Gain Score[a] (Mean)
Reading	Public School Students	563.75*	612.96	49.21
	3-year Former Recipients	544.76	618.62	73.86*
	2-year & 1-year Former Recipients	546.50	601.16*	54.66
Language	Public School Students	560.88*	606.96	46.08
	3-year Former Recipients	542.03	606.47	64.44*
	2-year & 1-year Former Recipients	554.01	589.25*	35.24
Mathematics	Public School Students	514.37*	591.46	77.09
	3-year Former Recipients	504.02	595.95	91.93*
	2-year & 1-year Former Recipients	506.06	576.96*	70.90
Total Achievement	Public School Students	546.34*	603.79	57.45
	3-year Former Recipients	530.27	607.01	76.74*
	2-year & 1-year Former Recipients	535.52	589.12*	53.60

Public School Students, N = 1017. 3-year Former Scholarship Recipients, N = 38. 2-year/1year Former Scholarship Recipients, N = 60.

a. Within the 1st grade, 3rd grade, and Gain Score columns for each measure of achievement, the mean identified with an asterisk (*) differs at a statistically significant level from the other two means, which do not differ at a statistically significant level.

Summary: Differential Exit Groups in Comparison to Four-year Continuous Scholarship Recipients and Public School Students

In general, students who exited the CSTP to attend public schools displayed lower achievement at each testing episode (from first grade through third grade) in comparison to (a) scholarship students who remained in the program over this time period and (b) their peers in public schools. During first grade, former scholarship recipients achieved at a lower level than both scholarship recipients who remained in the program through third grade and public school students. By the end of third grade, most of the recipients who had exited the program maintained this lower level of achievement in comparison to both continuing scholarship recipients and their peers in public schools. Consequently, most of the students in the sample who withdrew from the CSTP to attend public schools were among the lower achieving scholarship students prior to, at, and following the time they exited the program.

The exceptions to these general findings, however, are the former scholarship recipients who exited the program after kindergarten to attend public schools for three years. Although recipients who exited the program after kindergarten achieved at a lower level than did continuing scholarship recipients and public school students at the end of first grade, they eliminated this initial achievement difference and performed at a level similar to continuing scholarship recipients and public school students at the end of third grade. Furthermore, former recipients who exited after kindergarten also gained more on all of the achievement measures in comparison to other former scholarship recipients who remained in the program longer and exited after first or second grade. Consequently, former recipients who attended public schools for three years achieved at a higher level on all of the achievement measures at the end of third grade in comparison to other former recipients who spent more time in the program.

4.0 Summary and Conclusions

This section is devoted to a brief summary of the findings associated with each of the three questions that guided the evaluation during the 2000-2001 year, and to a discussion of conclusions that can be drawn from these findings. Several points should be made, however, prior to engaging in this discussion.

First, unless otherwise noted, the results are based primarily on a cohort of students who entered public or private school as first graders during the 1998-99 academic year. Data on these students, and reported herein, were collected most recently in the spring of 2001, when cohort students were completing third grade. To the extent that this cohort of students is representative of the broader population of scholarship and public school students, which it generally appears to be, the findings can appropriately be generalized to the broader program. However, because the characteristics or experiences of this cohort may be unique in some ways from those of other students, generalizations must be made cautiously.

Second, any examination of the impacts of an educational program is a complex and difficult endeavor. This complexity is increased when the program is relatively large, encompasses a large number of schools, families, and students, and must be examined over time. Any attempt to draw broad or highly general conclusions across a range of findings requires a degree of simplification that cannot convey important nuances in the program or in the results.

Third, the findings reported above and discussed below are only the most recent from an ongoing study of the Cleveland Scholarship and Tutoring Program (CSTP). As such, they represent only a snapshot of the program and the cohort of students as they were in May of 2001. Definitive, summative, or final conclusions about the program, schools, or students cannot and should not be drawn. Collectively, these conditions require that caution be exercised as findings are synthesized and conclusions drawn.

4.1 Question One: What are the characteristics of students who participate in the Cleveland Scholarship and Tutoring Program, and how do they compare with students who do not participate?

In general, students who were using a scholarship in third grade during 2000-01 were from families of somewhat higher income and proportionally more were Caucasian, Hispanic, and Multiracial than their public school peers. Both scholarship and public school students were nearly equally male and female.

Only 52.5% of third grade scholarship students were African-American and 33.8% were Caucasian. This is in contrast to public school students, nearly 74% of whom were African-American and only 23.1% of whom were Caucasian. Interestingly, however, the proportion of scholarship recipients who were Hispanic (8.2%) or Multiracial (3.9%) was more than twice that for public school students (3.4% Hispanic, 1.7% Multiracial). Furthermore, scholarship users tended to be from families of comparatively, though not substantially higher income than public school students (i.e., less likely to qualify for free or reduced lunch).

Students who were awarded a scholarship but chose to enroll in public schools (i.e., winner non-users) were of significantly lower income (i.e., more likely to qualify for free lunch) than any other group in the study. Current data do not address the causes of this situation. However, data collected through interviews of parents and guardians in an earlier phase of the evaluation (See Metcalf et al., 1999)[74] would suggest that these families may have been unable to enroll their children in schools of choice near their home. Because they are of extremely low income, they do not possess the resources necessary to pursue enrollment in other schools and, as a result, they elect to remain in public schools. Data were again collected from many of these families through telephone interviews in spring, 2002 and may help our understanding.

Mobility among scholarship (private) and public school students was roughly equal, with the majority of students in both settings having attended the same school from kindergarten through third grade.

Notably, non-applicant private school students demonstrated the fewest number of school changes from kindergarten to third grade (93.4% had not changed schools at all). Scholarship

[74] Metcalf, K.K., *Evaluation of the Cleveland Scholarship and Tutoring Program: 1996-1999*, Bloomington, Indiana; Indiana Center for Evaluation, 1999. All prior reports are available through the Ohio Department of Education or through the Indiana Center for Evaluation (www.indiana.edu/~iuice).

students and those in public schools also were likely to have remained in the same school during this period (approximately 80% across the groups), and most of the remaining students in these groups had changed schools only once.

> *The variables of student gender, race/ethnicity, income, and mobility, both individually and collectively, had only very weak relationships to the achievement of students in the cohort. Higher overall academic performance, however, was generally associated with being Caucasian, of higher estimated family income, less mobile, and female. This was true for both scholarship and public school students.*

Being Caucasian and from a family of higher income explained from 5.6% of total achievement variance in mathematics and social studies to 7.3% in science. In addition, student mobility was found to explain from 5% to 7% of variance in language and total scores at the end of second and third grade, with students who changed schools more frequently achieving at lower levels. Student sex emerged as the best independent predictor of achievement *change* across time (i.e., from early first grade to late third grade). Female students were associated with greater amounts of positive achievement score change (gain) in each of the four subtest areas than were male students.

It must be noted, however, that the relationship between these demographic and mobility variables and student achievement, although statistically significant in some cases, is remarkably small. The variables collectively accounted for less than 1% of the variance in achievement score change across the four testing episodes.

> *Differences in demographic characteristics between scholarship and public school students at third grade may be attributed to the unique characteristics of students who joined the cohort in particular years. However, there are few clear patterns of demographic difference across years.*

For example, the third grade scholarship cohort generally tends to have fewer minority students than do the public school groups. However, only the most recent entrants into the program (in third grade, 2000-2001) and the initial entrants (in kindergarten, 1997-1998) have significantly smaller proportions of minority students. The proportion of minority students who entered the CSTP as first graders (in 1998-1999) and second graders (in 1999-200) does not differ significantly from the proportion of minority non-applicants in public schools. In fact, during

these years, scholarship entrants are proportionally *more* likely to be minority than are applicant non-recipients and winner non-users.

Students in the cohort who entered the CSTP in more recent years are likely to have applied for but not received a scholarship in previous years, and a majority tends to have been enrolled in private schools prior to entering the program..

Based on limited available data, proportionally more students who joined the cohort in second or third grade came from private schools (68% of 111) than from public schools (32% of 111). Further, a majority of these more recent entrants applied for but did not receive a scholarship in previous years. Interestingly, however, former public and former private school students are about equally as likely to have applied for a scholarship in prior years. Approximately 58% of former public school students and 62% of former private school students who entered the current cohort of scholarship users applied for but did not receive a scholarship in one or more previous years.

The characteristics of students who accept their scholarship later each year (i.e., after the beginning of the academic year) differ dramatically from the characteristics of students who accept their scholarships prior to the beginning of school.

Supplemental analyses of data on scholarship students in grades 2, 3, 4, and 8 indicate that a majority of students enter the program from public schools (67.8%). Of these, students who are awarded and accept their scholarships prior to the beginning of the academic year are nearly identical to their public school peers in ethnicity or race. Only 21.8% of pre-year scholarship entrants were Caucasian. Further, 72.9% of these students had attended a public school in the year preceding their entry into the CSTP. In contrast, students who were awarded and accepted their scholarships after the beginning of the school year were about equally Caucasian or of minority status (49.3% and 51.7%, respectively). A slight majority of these students (55.7%) had attended a private school in the preceding year.

On the basis of these findings, it seems that the random lottery, held by the CSTP office each spring, generally serves to award scholarships to students who have been in public schools and who are very similar to their peers in Cleveland public schools. As the beginning of the academic year approaches and then passes, however, some students elect not to use scholarships

that they were awarded. These scholarships are then offered to other applicants to the program (who may not have been selected in the initial lottery). Students who, at this later point, are awarded and choose to use a scholarship tend to be Caucasian and to have been enrolled in private school previously.

> *Differences between the recent scholarship enrollees and public school students appear primarily to be a result of the characteristics of students who enter the program from private schools. Students who move into the program from public schools are nearly identical to their public school classmates.*

As the analyses discussed above indicate, recent scholarship recipients as a whole (i.e., all students who entered the program in second and third grade) are *less likely* to be minority students and *more likely* to be from families with higher incomes than either (a) recipients who entered the program in kindergarten and first grade or (b) their peers in public schools. However, this trend differs between students who entered the program from public and private schools.

Students who entered the CSTP in second and third grade from public schools were statistically equivalent to their classmates who remained in public schools. Approximately 80% of these students were minority and 87% were from families of lower income (i.e., eligible for free lunch). In contrast, scholarship recipients who entered the program in second grade and third grade after attending *private schools* in previous years differed significantly from their peers in public schools. Recent recipients who attended *private schools* prior to entering the CSTP were approximately 46% minority and only 54% were of lower income.

Furthermore, among the 72 recent scholarship recipients whose previous schools of enrollment are *unknown*, both the proportion of minority students (76%) and the proportion of lower income students (85%) were statistically equivalent to the proportion of minority and lower income students in the public schools. Unknown students represent those who do not appear in any of the three major data sets drawn upon in the present project (i.e., Cleveland Municipal School District records, CSTP office records, and the multi-year dataset maintained by the evaluation team). Thus, these unknown students are likely to have moved into the Cleveland area from another district.

Students who choose to exit the CSTP after one or more years are much more similar to their public school counterparts than they are to scholarship students who continue in the program.

Regardless of when former scholarship recipients withdrew from the program, they tended to be comprised of a greater number of minority students and students from families with lower incomes in comparison to scholarship recipients who continuously participated in the program from kindergarten through third grade. However, the demographic characteristics of former scholarship recipients did not differ significantly from their peers in public schools (applicant non-recipients, winner non-users, or non-applicants) on any of the variables we examined. That is, students who chose to exit the CSTP tend to be more demographically similar to public school students than they are to the scholarship recipients who have chosen to remain in the program through third grade, 2001. A majority of both former winners and public school students are minorities (86.6% and 79.9%, respectively), and students in both groups are likely to qualify for free lunch.

4.2 Question Two: What are the characteristics of the classrooms and teachers with whom scholarship students work in private schools, and how do they compare with the characteristics of classrooms and teachers in public schools?

Private school teachers have significantly more years of teaching experience than their public school counterparts and are more likely to have taught in the same school for a significantly longer period. However, public school teachers are more likely to be fully certified than private school teachers. No differences were found in the highest degree obtained by teachers or in class size between public and private schools.

Private school teachers who worked with scholarship students had an average of 14.4 years of professional experience and had taught in the same school for an average of nearly half of that time (7.3 years). Public school teachers also were experienced, but had only an average of 10.9 years of experience, roughly half of this (5.4 years) in their present school. A majority of the teachers who worked with cohort students were fully certified, but significantly more public school teachers than private school teachers were certified (88% and 96%, respectively). Private and public school teachers were proportionally equally likely to have completed some graduate

coursework. Class size was surprisingly similar between public and private schools. Students in both settings worked in classes of nearly 20 students, on average.

None of the teacher or classroom variables (singly or collectively) explained a significant portion of achievement change. Furthermore, although class size emerged as the single best predictor of student achievement at particular points in time, this relationship is extremely weak and positive.

Student achievement variance attributable to the teacher or classroom variables ranged from less than 2% (in reading and language) to just over 4% in mathematics. Further, only class size was found to be related to achievement growth (change over time), explaining from 1.6% of the variance in reading achievement growth over time to up to 3.1% of the variance in mathematics achievement. It is interesting, however, that the relationship between class size and student achievement growth is *positive*. In other words, students in *larger* classes were likely to experience *greater achievement growth* over time than those in smaller classes. The magnitude and directionality of this finding reflect the problems associated with attributing student achievement to class size without considering mediating variables.

The teachers and classrooms experienced by minority and non-minority students tended to differ, regardless of whether they were attending public or private schools.

Minority students attended significantly smaller classes than did non-minority students, regardless of whether they were scholarship recipients in private schools or public school students (22.6 students and 24.1 students respectively). Furthermore, minority students in both public and private schools tended to work with teachers who were somewhat more likely to have completed graduate coursework than non-minority students. Non-minority students were proportionally more likely to work with fully certified teachers (96%) than were minority students (92.7%), regardless of whether they attended a private or a public school. Although teacher experience was statistically the same for minority and non-minority students in private schools, minority students in public schools worked with significantly less experienced teachers (10.7 years) than non-minority students (13.2 years). No differences were found between minority and non-minority students on other classroom variables.

4.3 Question Three: What is the impact of participation in the Cleveland Scholarship and Tutoring Program on student academic achievement?

Before presenting conclusions associated with this question, it is worth reviewing the data upon which the analyses were based. Achievement data have been collected on the current cohort of students from the beginning of first grade in 1998 through the end of third grade in 2001. The standardized achievement test that is used for this purpose provided four measures of achievement through the end of second grade (Terra Nova Levels 10, 11, and 12). These included scores in reading, language, mathematics, and an aggregated total score. The third grade test (Level 13) adds measures in the areas of science and social studies.

The primary analyses focus on examining and comparing the achievement of students in each of five groups from the beginning of first grade through the most recent testing episode (end of third grade). Two of these consisted of students who were using a scholarship to attend private school: *four-year scholarship students*, who had used a scholarship continuously since kindergarten, and *three-year scholarship students*, who had begun using a scholarship in first grade. The three remaining groups attended public schools: *applicant non-recipients*, who had applied for a voucher in one or more years but had not received one, *winner non-users*, who had been awarded a voucher but chose not to use it; and *non-applicants*, students who attended public schools and had never applied for a scholarship. Thus, these analyses rely on scores in reading, language, mathematics, and a total score that are available throughout the period of the study. A secondary set of analyses compares the performance of scholarship and public school students in the areas of science and social studies. Because scores on these measures are available only from third grade, they cannot account for change over time, but they do provide additional indicators of potential areas of achievement difference among the student groups.

> *Scholarship students were achieving at much higher levels than their public school counterparts at the beginning of first grade, but these differences were eliminated by the end of first grade. Across the four measures available during this period (reading, language, mathematics, and total score), relative commonality was maintained across the groups through the end of second grade.*

Our previous report (Metcalf et al., 2001)[75] details the findings associated with student academic performance from the beginning of first grade through the end of second grade. The results of the present analyses, which include more accurate identification of students and imputation of missing data, are consistent with those earlier findings. Scholarship students (both those entering in kindergarten and those entering in first grade) were achieving at significantly higher levels in reading, language, mathematics, and total score than were students in any of the public school groups (including winner non-users, applicant non-recipients, and non-applicants). With the exception of reading, there were no significant differences between or among the scholarship and public school groups by the end of first grade. In reading, winner non-users were achieving at higher levels by the end of first grade than were scholarship students who had begun the program as kindergartners.

In general, the student groups remained relatively similar through the end of second grade. There were no differences among the groups in reading or total achievement. Interestingly, although students who entered the CSTP in kindergarten were achieving at higher levels than non-applicants or applicant non-recipients in language at this point, they were achieving at significantly lower levels in mathematics than non-applicants.

Across the measures of academic achievement, students in each group demonstrated significant growth across time. The greatest single period of growth was from the beginning to the end of first grade, particularly for public school students. During this period, students whose families had received a scholarship but had chosen not to use it exhibited consistently greater growth than did any other students, and scholarship students demonstrated relatively less growth. It must be noted, however, that statistical regression toward the mean may explain some of this difference in achievement growth.

During second grade, student academic achievement growth was largely equal across the areas of reading, mathematics, and total score. However, scholarship students who began in kindergarten demonstrated significantly greater growth in language achievement than did either non-applicants or applicant non-recipients.

[75] Metcalf, K.K., *Cleveland Scholarship and Tutoring Program Evaluation: 1998-2000 Technical Report*, Bloomington, Indiana; Indiana Center for Evaluation, 2001); available through the Ohio Department of Education or through the Indiana Center for Evaluation (www.indiana.edu/~iuice).

It should be noted that the highly divergent scores at the beginning of first grade may mask possible program effects on the achievement growth of scholarship students. Because the initially high performing students are disproportionately represented among three-year and four-year scholarship recipients, it is not possible to determine whether these students demonstrate less change across time because of factors specific to the CSTP, or simply because of the statistical effect of regression toward the mean. A similar issue arises when attempting to interpret the greater amount of achievement change displayed by the initially low performing public school comparison groups. Because regression toward the mean remains a viable alternative explanation, some caution should be exercised when interpreting any of the findings that are indicative of differential achievement change across time between two or more of the student groups.[76]

By the end of third grade, no consistent differences in achievement were found between or among the scholarship students and their public school peers on any of the four longitudinal achievement measures.

There were no differences in achievement between scholarship and public school students in reading or language by the end of third grade. Three-year scholarship recipients were achieving at significantly lower levels in mathematics than were any other group by this time, and at significantly lower levels than were four-year scholarship recipients in total score. Although each group improved significantly from the end of second to the end of third grade, the achievement growth both of three-year and of four-year scholarship recipients from the beginning of first grade through the end of third grade was less than the achievement growth of public school students. However, this is almost solely attributable to the higher initial performance of these groups and achievement growth of public school students during first grade. From the end of first grade, when the groups were performing at similar levels, through the end of third grade, achievement growth was roughly similar across all of the groups. However, three-year scholarship students tended to gain less during this period than other students in the areas of mathematics and total score.

[76] A discussion of regression toward the mean as a statistical artifact, and the potential for this effect to confound the current findings, is presented in Appendix C for interested readers.

There were no significant differences among the groups in science or social studies achievement at the end of third grade.

As noted earlier, achievement in science and social studies was measured only once and at the end of third grade. Thus, these third grade scores represent what must be considered baseline data on these measures. At the end of third grade, both three-year and four-year scholarship students achieved at generally equal levels on the science and social studies measures in comparison to their public school peers.

In general, the duration of time students spent in the CSTP was not related to achievement.

There were no consistent differences in third grade achievement between students who had participated in the program continuously since kindergarten, first grade, second grade, or third grade. Students' achievement in reading, language, mathematics, social studies, and total was similar across these groups of scholarship students. The single significant difference found between the scholarship groups at the end of third grade was in science achievement. Scholarship recipients who participated in the program since kindergarten achieved at a significantly higher level on the third grade science measure in comparison to recipients who entered the program as second graders.

Students who leave the program after one or more years to enroll in public schools tend to have been achieving at lower levels than other scholarship students. Furthermore, this lower level of academic achievement generally continues after they move to public schools.

In general, students who exited the CSTP to attend public schools displayed lower achievement at each testing episode – from first grade through third grade – in comparison both to scholarship students who remained in the program over this time period and to their peers in public schools. In other words, most of the students in the sample who withdrew from the CSTP to attend public schools were among the lower achieving scholarship students prior to, at, and following the time they exited the program.

Despite relatively low achievement at the beginning of first grade, former scholarship users who withdrew after kindergarten made considerable achievement gains in the public school system over the next three years and were performing at a level similar to their classmates in public schools and to scholarship users who remained in the program through third grade.

Unlike scholarship recipients who exited the program between first and third grade, recipients who exited after kindergarten to attend public schools for the next three years displayed substantial achievement gains and were performing at a level similar to continuing scholarship recipients and public school students by the end of third grade. Notably, at the end of third grade, this group of former scholarship recipients outperformed the other former recipients who spent more time in private schools.

For an array of reasons, some scholarship students and their families may discover that the private schools they selected do not suit their needs as well as the public school environment they left behind. Although purely speculation at this point, a poor student-environment "fit" reasonably could explain the lower performance of scholarship recipients who eventually exit the program. In addition, the transition from private to public schools may further inhibit the academic achievement of former scholarship recipients, such that the benefit of a better student-environment fit does not manifest immediately. Hence, the students in the sample who were the first to exit the CSTP may be the only group that has had sufficient time to realize the advantages of participating in an academic environment better suited to meet their needs.

Future evaluation activates will be tailored to empirically address this hypothesis. Not only will more recent former recipients be targeted to investigate whether a similar pattern of achievement gain eventually emerges, but other methodologies also will be utilized (e.g., surveys) to specifically investigate the issue of student-environment fit as a factor that may mediate satisfaction with, and the likelihood of withdrawal from, the CSTP.

4.4 Summary

The results that have been presented in the current report are drawn from work conducted during the period fall, 1998 through spring, 2001 that represents only a portion of a multi-year, longitudinal study of the voucher program in Cleveland. The primary focus of the work has been and remains a cohort of students who entered public or private schools as first graders in 1998 and who, in spring 2001, were at the end of third grade. During the period of the current report, data collection and analysis were guided by three questions associated with the characteristics of students, the characteristics of classrooms and teachers with whom them worked, and the impact of the Cleveland Scholarship and Tutoring Program on students' academic achievement.

Other researchers of publicly funded vouchers, including our own earlier work, have produced differing findings and have reached vastly different conclusions. We have speculated in other arenas that these divergent findings may be the result of differing methodological approaches to the study of voucher programs and to the cumulative effects of education (and thus of voucher programs) that studies conducted over only three or four years may not identify (see Metcalf et al., 1999; Metcalf & Beghetto, 2001; Metcalf & Legan, 2001). The present study addresses these issues in ways that, we believe, may prove useful in reconciling these seemingly disparate results.

In the earlier phase of our work in Cleveland, we found that scholarship students were somewhat more likely to be Caucasian and of much lower income than their public school counterparts. The current results offer a somewhat different picture. As in our earlier work, a greater proportion of the students in the third grade cohort are Caucasian, and fewer are African-American. However, the proportion of Hispanic and multiracial students in the third grade scholarship cohort is roughly twice that of the public schools. Thus, although the cohort is more Caucasian than the public schools, it is also more broadly diverse. Our current findings on family income of scholarship students contrast with our earlier findings. Scholarship students in the cohort are less likely to qualify for free lunch than their public school peers.

Students who enter the program after first grade tend to come from families who have applied for a voucher in prior years. There is also a slight tendency for them to have been enrolled in a private school before entering the CSTP. However, and surprisingly, the proportion of new

scholarship students whose families have previously applied for a scholarship is nearly identical for former public and former private school students. It appears that many families become interested in obtaining a scholarship early in their children's schooling, and they persist in their efforts to do so over multiple years.

More important, a clear pattern emerges in the characteristics of students who are awarded and accept their scholarship prior to the beginning of the academic year and those who do so after the beginning of school. Students who accept their scholarship prior to the beginning of school are nearly identical to their public school classmates in family income and race, and a majority is from public schools. However, a majority of students who accept their scholarship after the beginning of the school year are former private school students (likely already enrolled in the private school), are of higher income, and are less likely to be African-American.

As in previous years, we found both similarities and differences between public and private school classrooms and teachers. The vast majority of teachers in both types of school were fully certified, although the proportion was greater in public schools than in private. Private school teachers had more experience and had been in their current schools longer than public school teachers. Public school teachers were more likely to have completed some graduate coursework, but most teachers in both types of school had done so. Class sizes were slightly larger in private schools, but both public and private school classrooms contained a greater number of students than many would recommend (e.g., the Tennessee STAR report). Of these variables, only class size was related to student achievement and this relationship, surprisingly, was positive (i.e., larger classes were associated with higher academic achievement).

An interesting trend is found in the classrooms experienced by minority and non-minority students in the cohort we have studied. Minority students tend to be enrolled in smaller classes and with teachers who have more experience, whether they attend private or public schools. However, non-minority students are more likely to work with teachers who are fully certified. These findings, while significant, may be a result of unique characteristics of the third grade classrooms. Or, schools may attempt to secure more experienced teachers and smaller classes for minority students. Future years of the study will allow us to identify patterns that may exist.

The most recent results associated with student achievement do not reveal any significant impacts of participation on the CSTP on students' achievement. From the end of second grade, when the large initial differences between public schools students and scholarship students no longer existed, students in all of the groups we have studied demonstrated significant gains each year. And, across groups, the general extent of achievement growth was nearly equal through second and third grade. Though it is not significant in the data available to date, there is some evidence of a pattern of slightly greater annual achievement growth among students who have used a scholarship continuously since kindergarten. If this pattern continues, the achievement of this group of students may become noticeably higher than that of public school students. However, data over three to five additional years will be necessary to confirm or discount such a pattern.

As we have found in previous years, students who choose to leave the CSTP and enroll in public schools are achieving at lower levels than any other group of students in our study. This trend is found regardless of the grade at which they discontinue the program. The data available on these students over multiple years now also indicate that this pattern of comparatively low achievement continues for these students in their public schools. A notable exception to this trend is the group of students who left the program after kindergarten and who have attended public school consistently since that time. These students were achieving at roughly equal levels to other scholarship students at the beginning of first grade, and continue to do so. For other students who discontinue in the program, future data collection will attempt to identify potential explanations for their achievement.

4.5 Closing Thoughts

The recent ruling by the U.S. Supreme Court that the Cleveland Scholarship and Tutoring Program is constitutional has removed a cloud of doubt about the program's future that has loomed over it since it began. Many researchers and policy makers now speculate that other states will propose new voucher programs, and that existing programs, particularly Cleveland's, will now thrive. Even opponents of such programs have increased their calls for sound, empirical research on these programs that can inform the re-ignited policy debate.

The longitudinal study that is being conducted in Cleveland represents the longest running, and most comprehensive evaluation of a publicly funded voucher program that has ever been conducted. As such, it affords a unique opportunity to examine the characteristics, operation, and impacts of these programs over time. The data that have been collected thus far, and that are reported above, represent work done through spring of 2001. However, the evaluation continues.

At the present time, two key additional data collection activities have already taken place. Students in the cohort, who were enrolled in fourth grade during the 2001-02 year, were administered the standardized achievement test by the evaluation team in April of 2002. These tests are currently being scored and these data will be integrated into the existing data set. As a result, we will shortly have available an additional year of achievement data for these students. Further, plans are underway to conduct testing of the cohort in spring of 2003.

In addition, telephone interviews were conducted with random samples of parents and guardians of both public and private school students in Cleveland. These extensive interviews included a range of critically different groups of families, each of which possesses unique attributes which may influence their perceptions of schools and school choice. Sampling across grades 2, 4, and 8, interviews were conducted with families who:

1. Had been awarded and were using scholarships to send their children to private schools (representing what we have called *winner users),*
2. Had applied for, but were not awarded scholarships and had enrolled their children in public schools (representing what we have called *applicant non-recipients),*
3. Had applied for and were awarded scholarships, but who chose instead to enroll their children in public schools (representing what we have called *winner non-users*),
4. Had never applied for scholarship, and who had enrolled their children in public schools (what we have called *non-applicants),* and importantly,
5. Had chosen to enroll their children in charter schools (referred to as community schools in Ohio).

Data drawn from these interviews will provide insights into the criteria by which families make decisions about their children's education. The inclusion of families of charter school students and the explicit inclusion of winner non-users represents a new approach to this type of study of school choice.

The recent Supreme Court ruling and the renewed importance of rigorous research on this issue raise a number of new questions. For example, will applications to the Scholarship and Tutoring Program in Cleveland increase dramatically because the stability and future of the program are now secure? Anecdotal evidence from past years suggests that some families elected not to pursue or use a scholarship because of the program's uncertain legal status. Relatedly, will private schools respond to the ruling by creating more classrooms for scholarship students, or will new private schools open or elect to participate in the program? Will the characteristics of families who apply for scholarships change? It is not unreasonable to believe that previous uncertainty about the program may have led to a disproportionate number of applicants to be families whose children already attended private schools. These families would not have to worry about losing their scholarship if the program were discontinued. In contrast, many families whose children were in public schools may have been interested in a scholarship for private school enrollment, but they chose not to risk the disruption of moving their children back into public schools if the program was discontinued.

What are the long-term impacts of vouchers on schools, families, or students? Answers to this question and a range of others can be found. The ongoing work in Cleveland is one step toward this goal, as is the work of other researchers in other parts of the country. However, it will take time, careful study, and patience to obtain the evidence that will be required by policy makers and families as the future of school choice unfolds.

5.0 Future Evaluation Activities

The current evaluation is longitudinal and it is anticipated that the evaluation team will continue collecting data on students, their families, and their schools for several more years. Although the exact focus of future evaluation activities will be determined by information needs and resource availability, we anticipate they will include the following:

- Classroom and teacher data will be collected as in the past, but will include information about the amount of time spent on core subjects and resources (e.g. textbooks, workbooks, manipulatives, etc.) available for teaching these subjects. These additional data were collected by proctors during the fourth grade test administration in April, 2002, and will be included in the next report.

- Achievement data will continue to be collected annually on the target students. The Terra Nova level 14 was administered to fourth graders in April, 2002 and will be included in the next report. The next level of the Terra Nova (level 15) is anticipated to be administered in the spring of 2003 to fifth graders.

- Telephone interviews with randomly selected parents from our sample are to be conducted spring, 2002, the results of which will be included in the next report. Specifically, interviewees will be selected from pools of public school non-applicants, Public school applicant non-recipients, scholarship winner-users, scholarship winner non-users, as well as former scholarship students. These interviews parallel those conducted during the first phase of the evaluation in spring of 1999, but they are intended also to examine issues related to the legal uncertainty of the program and its impact on parents, parents' reasons for pursuing or not pursuing a scholarship, and parent perceptions of the type and quality of instruction provided to their children, to name a few.

- Written surveys of teachers, school level administrators, and students also will be conducted during 2002-2003 school year to gather information on perceptions and awareness of the Cleveland Scholarship and Tutoring Program, and these surveys will be supplemented with some focus groups.

- Additionally, there is a desire to obtain information related to the impact of the CSTP on the public and private schools in Cleveland. If resources allow, an examination of the financial, procedural, operational, and political impacts of the program could provide information enabling us to begin to understand how the program may have influenced the educational context in Cleveland.

Appendix A

Multiple Imputation Procedure

Because of student transience, inconsistencies with school records, absences during the testing episodes, and other inevitable but unintended factors that could not be controlled by the researchers, some students were not able to be located or were unavailable during one or more of the data collection periods. Consequently, out of the 2428 students in the sample who were classified as members of one of the five target groups as of spring, 2001, approximately 60% (1430) were missing at least one out of the four possible scores on at least one of the achievement measures. In other words, a complete set of achievement scores on the reading, language, and mathematics measure from all four testing episodes was obtained from only 40% of the students in the sample. Table 60 displays the number and percentage of students in each student group, and overall, with complete data from all of the testing episodes. Furthermore, Table 61 separately displays the amount of missing and complete data on each achievement measure across the four testing episodes.

As these tables demonstrate, although the rates of missing data on each individual achievement measure from a given testing episode are relatively low, substantially fewer students have complete achievement data across all four testing episodes. Because case-deletion procedures might bias the results of the statistical analyses if the 1430 students missing one or more achievement datum are unrepresentative of the entire sample (or of their respective scholarship status groups), it was necessary to utilize procedures to address the issue of missing data.

Missing Data Estimation and Replacement Procedure

In general, multiple imputation is a statistical simulation procedure (using a Markov chain Monte Carlo technique) in which each missing value is replaced by $m > 1$ plausible values, such that $m > 1$ separate and complete data sets are created and subsequently analyzed individually using standard statistical software. The results of the $m > 1$ separate analyses are then combined using algorithms developed by Little and Rubin (1987).[77] The variability among the $m > 1$ values that

[77] Little, R.J.A, & Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* Wiley & Sons, New York.

replace each missing datum captures the error involved with predicting unknown values from observed values. Each simulation of a missing achievement score includes both a regression prediction and a residual error term randomly drawn from a probability distribution.

Table 60. 2001 Sample: Number of Complete and Incomplete Student Achievement Data Records across the Four Testing Episodes

Student Group Membership as of Third Grade (Spring, 2001)[a]	Total in 2001 Sample	Observed Data Complete Cases Across Four Testing Episodes[b]		Missing Data Incomplete Cases Across Four Testing Episodes[c]	
	N	N	Percent Complete	N	Percent Incomplete
Scholarship Recipients					
1 year Scholarship Recipients	111	34	30.6%	77	**69.4%**
2 year Scholarship Recipients	79	21	26.6%	58	**73.4%**
3 year Scholarship Recipients	120	78	65.0%	42	**35.0%**
4 year Scholarship Recipients	361	231	64.0%	130	**36.0%**
Public School Comparison Groups					
Former Scholarship Recipients	112	33	29.5%	79	**70.5%**
Scholarship Winner Non-users	95	33	34.7%	62	**65.3%**
Applicant Non-recipients	355	173	48.7%	182	**51.3%**
Non-applicants	1195	395	33.1%	800	**66.9%**
Total N =	**2428**	**998**	**41.1%**	**1430**	**58.9%**

a. The scholarship status codes were assigned during the third grade testing episode (spring 2001) based on the trumping order classification system described above

b. The number and percentage of students in each scholarship status group with complete reading, mathematics, language and total achievement scores from each of the four testing episodes

c. The number and percentage of students in each scholarship status group who are missing at least one achievement score from one of the achievement measures

A multiple imputation procedure was used in the present investigation because simply substituting regression predictions for the missing data points fails to account for the uncertainty (i.e., error) of predicting missing data, which could result in: (a) artificially inflated correlations between the observed and simulated values; (b) artificially small standard errors; (c) p-values that are too small; and (d) Type I error rates that may be inflated. That is, when using simple linear regression techniques to estimate missing data, known as regression imputation, the predicted values that replace the missing data fall exactly on the regression line (or plane). This technique incorrectly assumes that the missing data can be perfectly predicted from the observed data without any error. Multiple imputation, however, incorporates an estimate of the error and

uncertainty involved in replacing missing scores with estimates of what the unknown values might have been. According to Graham and Schafer (1999), "Multiple imputations reflect the uncertainty of missing-data prediction by restoring the natural levels of variability in the imputed data. This variability comes from two sources: first, the residual variation of [observed values] above and below the regression line; and second, the variation in estimating the regression line itself (pp. 5-6)[78]

Table 61. 2001 Sample: Amount of Missing Data on each Achievement Measure (Reading, Language, Mathematics, and Total) across the Four Testing Episodes

Achievement Measure & Testing Episode	Valid – Observed Cases		Missing Cases		Total in 2001 Sample
	N	Percent	N	Percent	N
Reading					
Fall 1st Grade (1998)	1598	65.8%	830	**34.2%**	2428
Spring 1st Grade (1999)	1564	64.4%	864	**35.6%**	2428
Spring 2nd Grade (2000)	1627	67.0%	801	**33.0%**	2428
Spring 3rd Grade (2001)	2257	93.0%	171	**7.0%**	2428
Language					
Fall 1st Grade (1998)	1596	65.7%	832	**34.3%**	2428
Spring 1st Grade (1999)	1563	64.4%	865	**35.6%**	2428
Spring 2nd Grade (2000)	1629	67.1%	799	**32.9%**	2428
Spring 3rd Grade (2001)	2254	92.8%	174	**7.2%**	2428
Mathematics					
Fall 1st Grade (1998)	1604	66.1%	824	**33.9%**	2428
Spring 1st Grade (1999)	1586	65.3%	842	**34.7%**	2428
Spring 2nd Grade (2000)	1649	67.9%	779	**32.1%**	2428
Spring 3rd Grade (2001)	2255	92.9%	173	**7.1%**	2428
Total Achievement					
Fall 1st Grade (1998)	1551	63.9%	877	**36.1%**	2428
Spring 1st Grade (1999)	1505	62.0%	923	**38.0%**	2428
Spring 2nd Grade (2000)	1579	65.0%	849	**35.0%**	2428
Spring 3rd Grade (2001)	2165	89.2%	263	**10.8%**	2428
Science					
Spring 3rd Grade (2001)	2246	92.5%	182	**7.5%**	2428
Social Studies					
Spring 3rd Grade (2001)	2245	92.5%	183	**7.5%**	2428

[78] Graham, J.W., & Schafer, J.L. (1999). On the performance of multiple imputation for multivariate data with small sample sizes. In R.H. Hoyle (Ed.), *Statistical Strategies for Small Sample Research.* (pp. 1-20). Sage Publications, Thousand Oaks, CA.

Based on the recommendations presented by Little and Rubin (1987), ten separate data sets were imputed for the present achievement score analyses, such that each missing datum was simulated ten times. Because of the number and complexity of the mixed-design analyses planned for this project, the ten data sets were combined prior to conducting the analyses (as opposed to analyzing each simulated data set separately and then combining a large number of parameters across ten data sets via Little and Rubin's algorithms). By doing so, however, some of the benefit of multiply imputing the missing data was sacrificed. That is, the between-imputation variability, which captures a portion of the uncertainty involved with predicting unknown values from observed, was not statistically accounted for. Instead, the simulated achievement scores were averaged across the ten imputed data sets, and the average imputed values were merged with the observed values to create a single complete data set for the analyses. Preliminary analyses of variance indicated that the pattern of achievement on each measure (reading, language, mathematics, and total achievement) across the four testing episodes did not differ significantly among the ten imputed data sets, which supports the averaging approach used to combine the ten imputations. Although the variability among the simulated scores was lost by averaging across the ten imputed data sets, the benefit of originally replacing each missing score with a randomly selected value plus a residual error term from a probability distribution was preserved.

Creating the Imputation Model: Selective Case Inclusion

For a given variable, comprised of missing and observed values, multiple imputation replaces the missing data based on the relationships between the observed data on the target variable and the observed data on all of the other variables selected as relevant predictors of the missing data on the target variable. In longitudinal designs, multiple scores are obtained from each student across time. Therefore, the best predictors of a given student's missing achievement score(s) from one or more testing episodes likely are his or her observed achievement scores from other testing episodes. For example, for students missing achievement scores from the late second grade (spring, 2000) testing episode, the multiple imputation model should included achievement scores from the other testing episodes as predictors of the missing second grade scores. By doing so, better (i.e., more accurate and precise) predictions of missing achievement scores can be obtained for a particular individual when many, as opposed to few, achievement scores are present in the data set for that student and the other students in the sample.

To ensure that an adequate amount of achievement data was available from each student in the sample for inclusion in the multiple imputation model, students in each group with missing data from more than two testing episodes were excluded from the 2001 imputed sample. Because many of the scholarship recipients who entered the program after first grade were missing achievement data from the first two testing episodes (i.e., two-year and one-year scholarship recipients), these students were not included in the imputed data set. That is, missing achievement scores for one-year and two year scholarship recipients were *not* imputed retrospectively based on only the two scores obtained on each achievement measure during the second and third grade testing episodes. Rather, analyses addressing the question of differential entry into the CSTP included only the actual data obtained from these groups (i.e., only one-year and two-year scholarship recipients with complete achievement data across the second and third grade testing episodes were included in the analyses).

To preserve any differences in achievement between the student groups, each group was imputed separately and the resulting files were merged following the multiple imputation process. For instance, because the academic achievement of scholarship recipients and public school non-applicants may differ across the four testing episodes, the unique interrelationships among the achievement score variables for scholarship recipients had to be preserved by imputing their missing values separately from the public school non-applicants.

Appendix B

Statistics for Pairwise Comparisons

(F-statistic and p-value Tables)

Question 1: Student Demographic Characteristics

Measure:	Testing Episode / Interval		All scholarship v. Public school groups
Meal Code	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic p-value**
All scholarship recipients	>	Public applicant non-recipients	F(1, 1505) = 28.00, p < .001
All scholarship recipients	>	Public non-applicants	F(1, 1505) = 37.24, p < .001
All scholarship recipients	>	Public scholarship winner non-users	F(1, 1505) = 32.13, p < .001
Public applicant non-recipients	>	Public winner non-users	F(1, 1505) = 7.32, p < .001
Public non-applicants	>	Public winner non-users	F(1, 1505) = 7.14, p < .001

Measure:	Testing Episode / Interval		4 year scholarship recipients v. 1 year scholarship recipients
Sex (1=Male, 2=Female)	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic p-value**
4 year scholarship recipients	>	1 year scholarship recipients	F(1, 609) = 7.63, p < .008

Measure:	Testing Episode / Interval		1 year scholarship recipients v. 2, 3, and 4 year scholarship recipients
Meal Code	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic p-value**
1 year scholarship recipients	>	4 year scholarship recipients	F(1, 656) = 19.34, p < .001
1 year scholarship recipients	>	3 year scholarship recipients	F(1, 656) = 16.72, p < .001
1 year scholarship recipients	>	2 year scholarship recipients	F(1, 656) = 21.23, p < .001

Measure:	Testing Episode / Interval		1 year scholarship recipients v. Public school groups
Minority status (non-minority=1, minority =2)	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic p-value**
Public applicant non-recipients	>	1 year scholarship recipients	F(1, 1109) = 22.53, p < .001
Public non-applicants	>	1 year scholarship recipients	F(1, 1109) = 15.30, p < .001
Public scholarship winner non-users	>	1 year scholarship recipients	F(1, 1109) = 28.80, p < .001

Measure:	Testing Episode / Interval		1 year scholarship recipients v. Public school groups
Meal code	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic p-value**
1 year scholarship recipients	>	Public applicant non-recipients	F(1, 955) = , p <
1 year scholarship recipients	>	Public non-applicants	F(1, 955) = , p <
1 year scholarship recipients	>	Public scholarship winner non-users	F(1, 955) = , p <

Measure:	Testing Episode / Interval		Public applicant non-recipients v. 1 year scholarship recipients
Sex (1=Male, 2=Female)	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic p-value**
Public applicant non-recipients	>	1 year scholarship recipients	F(1, 1092) = , p <

Measure:	Testing Episode / Interval	2 year scholarship recipients v. Public scholarship winner non-users
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
2 year scholarship recipients	>	Public scholarship winner non-users	F(1, 923) = 8.93,	p = .003

Measure:	Testing Episode / Interval	Public non-applicants, Public applicant non-recipients v. Public scholarship winner non-users
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public non-applicants	>	Public scholarship winner non-users	F(1, 923) = 8.07,	p = .005
Public applicant non-recipients	>	Public scholarship winner non-users	F(1, 923) = 8.27,	p = .004

Measure:	Testing Episode / Interval	Public school groups v. 2 year scholarship recipients
Minority status (1=non-minority, 2=minority)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year scholarship recipients	F(1, 1080) = 6.85,	p = .009
Public scholarship winner non-users	>	2 year scholarship recipients	F(1, 1080) = 14.30,	p < .001
Public scholarship winner non-users	>	Public non-applicants	F(1, 1080) = 9.88,	p = .002

Measure:	Testing Episode / Interval	3 year scholarship recipients v. Public school groups
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year scholarship recipients	>	Public applicant non-recipients	F(1, 962) = 6.04,	p = .014
3 year scholarship recipients	>	Public non-applicants	F(1, 962) = 7.24,	p = .007
3 year scholarship recipients	>	Public scholarship winner non-users	F(1, 962) = 17.87,	p < .001

Measure:	Testing Episode / Interval	3 year public school group comparison
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	Public scholarship winner non-users	F(1, 962) = 8.07,	p = .005
Public non-applicants	>	Public scholarship winner non-users	F(1, 962) = 7.88,	p = .005

Measure:	Testing Episode / Interval	3 year scholarship recipients v. Public school groups
Minority status (1=non-minority, 2=minority)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year scholarship recipients	F(1, 1120) = 8.04,	p = .005
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 1080) = 15.69,	p < .001

Measure:	Testing Episode / Interval	Public school group comparison
Minority status (1=non-minority, 2=minority)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public scholarship winner non-users	>	Public non-applicants	F(1, 1120) = 9.76,	p = .002

Measure:	Testing Episode / Interval	4 year scholarship recipients v. Public school groups
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	Pubic applicant non-recipients	F(1, 1200) = 18.17,	p < .001
4 year scholarship recipients	>	Public non-applicants	F(1, 1200) = 23.18,	p < .001
4 year scholarship recipients	>	Public scholarship winner non-users	F(1, 1200) = 27.93,	p < .001

Measure:	Testing Episode / Interval	4 year public school comparison
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	Public scholarship winner non-users	F(1, 1200) = 7.71, p = .006	
Public non-applicants	>	Public scholarship winner non-users	F(1, 1200) = 7.53, p = .006	

Measure:	Testing Episode / Interval	4 year scholarship recipients v. Public school groups
Minority status (1=non-minority, 2=minority)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	4 year scholarship recipients	F(1, 1362) = 24.62, p < .001	
Public non-applicants	>	4 year scholarship recipients	F(1, 1362) = 14.37, p < .001	
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 1362) = 26.03, p < .001	

Measure:	Testing Episode / Interval	4 year public school comparison
Minority status (1=non-minority, 2=minority)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public scholarship winner non-users	>	Public non-applicants	F(1, 1362) = 9.15, p = .003	

Measure:	Testing Episode / Interval	4 year scholarship recipients v. Former scholarship recipients
Meal code	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 406) = 6.74, p < .017	

Measure:	Testing Episode / Interval	4 year scholarship recipients v. Former scholarship recipients
Minority status (1=non-minority, 2=minority)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 455) = 7.54, p < .01	
2 year former recipients	>	4 year scholarship recipients	F(1, 455) = 12.11, p < .008	

Question 2: Classroom and Teacher Characteristics

Measure:	Testing Episode / Interval	Private schools v. Public schools
Teacher's years of experience	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Private school teachers	>	Public school teachers	(1, 312) = 5.49, p = .02	

Measure:	Testing Episode / Interval	Minority status comparison
Class size	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Non-minority students	>	Minority students	F(1, 1460) = 22.65, p < .001	

Measure:	Testing Episode / Interval	Minority scholarship recipient comparison
Class size	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Non-minority scholarship recipients	>	Minority scholarship recipients	F(1, 1460) = 22.33, p < .001	

Measure:	Testing Episode / Interval	Minority public non-applicant comparison
Class size	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Non-minority public non-applicants	>	Minority public non-applicants	F(1, 1460) = 7.64, p = .006	

Measure:	Testing Episode / Interval	Scholarship recipients v. Public school groups
Teacher's highest degree	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	Scholarship recipients	F(1, 1437) = 17.06, p < .001	
Public non-applicants	>	Scholarship recipients	F(1, 1437) = 10.13, p = .002	

Measure:	Testing Episode / Interval	Scholarship recipients v. Public school groups
Teacher's years of experience	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Scholarship recipients	>	Public non-applicants	(1, 1460) = 14.25, p < .001	
Public applicant non-recipients	>	Public non-applicants	F(1, 1460) = 6.90, p = .009	

Measure:	Testing Episode / Interval	Minority scholarship recipients v. Minority public school groups
Teacher's years of experience	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Minority scholarship recipients	>	Minority applicant non-recipients	(1, 1460) = 11.06, p = .001	
Minority scholarship recipients	>	Minority non-applicants	F(1,1460) = 26.48, p < .001	

Measure:	Testing Episode / Interval	Scholarship recipients v. Public school groups
Teacher certification (1=no, 2=yes)	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	Scholarship recipients	F(1, 1449) = 15.49, p < .001	
Public non-applicants	>	Scholarship recipients	F(1, 1449) = 28.73, p <. 001	

Question 3: Student Academic Achievement

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Reading achievement	Early 1st	

Groups			F-statistic	p-value
3 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 27.52, p < .001	
4 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 37.30, p < .001	
3 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 9.15, p = .003	
4 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 7.86, p = .005	

Measure:	Testing Episode / Interval	Public school group comparison
Reading achievement	Early 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	Public non-applicants	F(1, 4059) = 10.68, p = .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Reading achievement	Late 1st	

Groups			F-statistic	p-value
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 4059) = 8.37, p = .004	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Reading Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	4 year scholarship recipients	F(1, 4059) = 12.63, p < .001	
Public non-applicants	>	4 year scholarship recipients	(1, 4059) = 26.63, p < .001	
Public scholarship winner non-users	>	4 year scholarship recipients	(1, 4059) = 9.17, p = .003	
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 10.64, p = .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Reading Achievement	Early 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year scholarship recipients	F(1, 4059) = 11.69, p < .001	
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 23.37, p < .001	
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 4059) = 6.59, p = .01	
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 12.99, p < .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Language Achievement	Early 1st	

Groups			F-statistic	p-value
3 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 13.70, p < .001	
3 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 17.86, p < .001	
3 year scholarship recipients	>	Public scholarship winner non-users	F(1, 4059) = 19.14, p < .001	
4 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 20.13, p < .001	
4 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 28.60, p < .001	
4 year scholarship recipients	>	Public scholarship winner non-users	F(1, 4059) = 21.36, p < .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Language Achievement	Late 2nd	

Groups			F-statistic	p-value
4 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 11.07, p < .001	
4 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 9.13, p = .003	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Language Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	4 year scholarship recipients	F(1, 4059) = 12.31, p < .001	
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 16.94, p < .001	
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 4059) = 13.64, p < .001	
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 4059) = 7.64, p = .006	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Language Achievement	Late 1st to Late 2nd	

Groups			F-statistic	p-value
4 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 7.23, p = .007	
4 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 6.10, p = .014	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Language Achievement	Early 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic p-value
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 7.01, p = 008
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 6.10, p = .014

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Math Achievement	Early 1st	

Groups			F-statistic p-value
3 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 16.70, p < .001
3 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 13.32, p < .001
3 year scholarship recipients	>	Public scholarship winner non-users	F(1, 4059) = 9.19, p = .002
4 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 31.12, p < .001
4 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 25.64, p < .001
4 year scholarship recipients	>	Public scholarship winner non-users	F(1, 4059) = 11.33, p < .001

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Math Achievement	Late 2nd	

Groups			F-statistic p-value
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 9.81, p = .002

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Math Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic p-value
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 4059) = 12.59, p < .001
Public applicant non-recipients	>	3 year scholarship recipients	F(1, 4059) = 14.79, p < .001
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 22.54, p < .001
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 4059) = 12.61, p < .001

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Math Achievement	Early 1st to Late 1st	

Groups			F-statistic p-value
Public applicant non-recipients	>	4 year scholarship recipients	F(1, 4059) = 9.51, p = .002
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 14.86, p < .001
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 4059) = 7.65, p = .006
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 7.33, p = .007

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Math Achievement	Early 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic p-value
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 4059) = 7.04, p = .008
Public applicant non-recipients	>	3 year scholarship recipients	(1, 4059) = 31.46, p < .001
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 35.25, p < .001
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 4059) = 21.66, p < .001
Public applicant non-recipients	>	4 year scholarship recipients	F(1, 4059) = 17.65, p < .001
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 22.10, p < .001
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 4059) = 10.12, p = .002

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Math Achievement	Late 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 4059) = 6.62, p = .01	
Public applicant non-recipients	>	3 year scholarship recipients	F(1, 4059) = 11.46, p < .001	
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 10.44, p = .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Total Achievement	Early 1st	

Groups			F-statistic	p-value
3 year scholarship recipients	>	Public applicant non-recipients	F(1, 4059) = 24.79, p < .001	
3 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 37.75, p < .001	
3 year scholarship recipients	>	Public scholarship winner non-users	F(1, 4059) = 18.68, p < .001	
4 year scholarship recipients	>	Public applicant non-recipients	1, 4059) = 35.55, p < .001	
4 year scholarship recipients	>	Public non-applicants	F(1, 4059) = 60.77, p < .001	
4 year scholarship recipients	>	Public scholarship winner non-users	F(1, 4059) = 18.92, p < .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship comparison
Total Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 4059) = 8.84, p = .003	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Total Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year scholarship recipients	F(1, 4059) = 7.91, p = .005	
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 13.63, p < .001	
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 4059) = 11.80, p = .001	
Public applicant non-recipients	>	4 year scholarship recipients	(1, 4059) = 23.19, p < .001	
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 39.01, p <.001	
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 4059) = 19.92, p < .001	

Measure:	Testing Episode / Interval	3 and 4 year scholarship recipients v. Public school groups
Total Achievement	Early 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year scholarship recipients	F(1, 4059) = 25.74, p < .001	
Public non-applicants	>	3 year scholarship recipients	F(1, 4059) = 37.85, p < .001	
Public scholarship winner non-users	>	3 year scholarship recipients	F(1, 4059) = 19.16, p < .001	
Public applicant non-recipients	>	4 year scholarship recipients	F(1, 4059) = 11.49, p < .001	
Public non-applicants	>	4 year scholarship recipients	F(1, 4059) = 24.73, p < .001	
Public scholarship winner non-users	>	4 year scholarship recipients	F(1, 4059) = 8.02, p = .005	
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 4059) = 7.19, p = .007	

Measure:	Testing Episode / Interval	Scholarship recipient comparison
Total Achievement	Late 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 4059) = 10.33. p = .001	

Differential Entry into the CSTP and Academic Achievement

Measure:	Testing Episode / Interval	Scholarship recipient group comparison
Reading Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year scholarship recipients	F(1, 528) = 6.10, p = .014	
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 528) = 9.49, p = .002	
1 year scholarship recipients	>	2 year scholarship recipients	F(1, 528) = 9.02, p = .003	
1 year scholarship recipients	>	3 year scholarship recipients	F(1, 528) = 8.94, p = .003	

Measure:	Testing Episode / Interval	Scholarship recipient group comparison
Reading Achievement	Late 2nd to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
1 year scholarship recipients	>	2 year scholarship recipients	F(1, 528) = 4.87, p = .028	

Measure:	Testing Episode / Interval	Scholarship recipient group comparison
Total Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year scholarship recipients	F(1, 525) = 14.20, p < .001	
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 525) = 15.94, p < .001	
1 year scholarship recipients	>	2 year scholarship recipients	F(1, 525) = 15.28, p < .001	
1 year scholarship recipients	>	3 year scholarship recipients	F(1, 525) = 12.31, p < .001	

Measure:	Testing Episode / Interval	Scholarship recipient group comparison
Total Achievement	Late 2nd to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
1 year scholarship recipients	>	3 year scholarship recipients	F(1, 525) = 5.71, p = .017	
4 year scholarship recipients	>	3 year scholarship recipients	F(1, 525) = 5.38, p = .020	

Differential Exit from the CSTP Academic Achievement

Measure:	Testing Episode / Interval	Scholarship recipients v. Former scholarship recipients
Reading Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 455) = 7.69, p = .006	
4 year scholarship recipients	>	2 year former recipients	F(1, 455) = 13.81, p < .001	

Measure:	Testing Episode / Interval	Scholarship recipients v. Former scholarship recipients
Language Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 455) = 11.44, p < .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 455) = 37.60, p < .001	

Measure:	Testing Episode / Interval	Scholarship recipients v. Former scholarship recipients
Math Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 455) = 10.54, p = .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 455) = 17.11, p < .001	

Measure:	Testing Episode / Interval	Scholarship recipients v. Former scholarship recipients
Total Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 455) = 12.17, p < .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 455) = 26.34, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Language Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 1109) = 8.07, p = .005	
Public applicant non-recipients	>	2 year former recipients	F(1, 1109) = 21.25, p < .001	
Public non-applicants	>	2 year former recipients	F(1, 1109) = 23.16, p < .001	
Public scholarship winner non-users	>	2 year former recipients	F(1, 1109) = 12.88, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Math Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 1109) = 10.46, p = .001	
Public applicant non-recipients	>	2 year former recipients	F(1, 1109) = 10.24, p = .001	
Public non-applicants	>	2 year former recipients	F(1, 1109) = 13.99, p < .001	
Public scholarship winner non-users	>	2 year former recipients	F(1, 1109) = 8.42, p = .004	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Total Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	2 year former recipients	F(1, 1109) = 11.50, p < .001	
Public applicant non-recipients	>	2 year former recipients	F(1, 1109) = 16.06, p < .001	
Public non-applicants	>	2 year former recipients	F(1, 1109) = 18.29, p < .001	
Public scholarship winner non-users	>	2 year former recipients	F(1, 1109) = 11.44, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Reading achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 1109) = 9.11, p = .003	
Public non-applicants	>	2 year former recipients	F(1, 1109) = 8.98, p = .003	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Reading Achievement	All	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 19.13, p < .001	
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 6.78, p = .01	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	All	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 25.98, p < .001	
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 12.66, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	Early 1st	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 22.08,	p < .001
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 11.74,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	Late 1st	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 12.66,	p < .001
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 31.39,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	Late 2nd	

Groups			F-statistic	p-value
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 10.27,	p = .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 15.68,	p < .001
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 13.32,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 66.76,	p < .001
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 10.01,	p = .002
3 year former recipients	>	2 year former recipients	F(1, 1365) = 20.31,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	Late 2nd to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 8.86,	p = .003
3 year former recipients	>	2 year former recipients	F(1, 1365) = 8.03,	p = .005

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Language Achievement	Late 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 6.50,	p = .01
3 year former recipients	>	1 year former recipients	F(1, 1365) = 6.58,	p = .01
3 year former recipients	>	2 year former recipients	F(1, 1365) = 18.65,	p < .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 10.64,	p = .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	All	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 8.70,	p = .003
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 20.89,	p < .001
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 6.56,	p = .01

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Early 1st	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 10.30, p = .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 16.67, p < .001	
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 41.95, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Late 1st	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 11.79, p < .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 6.92, p = .009	
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 16.95, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Late 2nd	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 11.12, p < .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 25.45, p < .001	
3 year former recipients	>	1 year former recipients	F(1, 1365) = 7.43, p = .007	
3 year former recipients	>	2 year former recipients	F(1, 1365) = 13.32, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 42.57, p < .001	
3 year former recipients	>	2 year former recipients	F(1, 1365) = 26.23, p <.001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Late 1st to Late 2nd	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 8.15, p = .004	
3 year former recipients	>	2 year former recipients	F(1, 1365) = 11.60, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Early 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 23.84, p < .001	
3 year former recipients	>	2 year former recipients	F(1, 1365) = 24.46, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Math Achievement	Late 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 10.33, p = .001	
3 year former recipients	>	2 year former recipients	F(1, 1365) = 19.76, p < .001	
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 7.58, p = .006	
1 year former recipients	>	2 year former recipients	F(1, 1365) = 6.93, p = .009	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	All	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 14.67,	p < .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 45.30,	p < .001
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 14.02,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Early 1st	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 11.49,	p < .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 38.53,	p < .001
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 28.70,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Late 1st	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 15.98,	p < .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 20.97,	p <.001
4 year scholarship recipients	>	3 year former recipients	F(1, 1365) = 45.57,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Late 2nd	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 22.59,	p < .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 40.02,	p < .001
3 year former recipients	>	1 year former recipients	F(1, 1365) = 6.93,	p = .009
3 year former recipients	>	2 year former recipients	F(1, 1365) = 9.27,	p = .002

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
4 year scholarship recipients	>	1 year former recipients	F(1, 1365) = 10.13,	p = .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 96.25,	p < .001
1 year former recipients	>	2 year former recipients	F(1, 1365) = 10.89,	p < .001
3 year former recipients	>	2 year former recipients	F(1, 1365) = 44.47,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Late 1st to Late 2nd	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 10.64,	p = .001
3 year former recipients	>	2 year former recipients	F(1, 1365) = 11.32,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Early 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 10.71,	p = .001
3 year former recipients	>	2 year former recipients	F(1, 1365) = 18.83,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Scholarship recipients
Total Achievement	Late 1st to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
3 year former recipients	>	4 year scholarship recipients	F(1, 1365) = 18.13,	p < .001
3 year former recipients	>	2 year former recipients	F(1, 1365) = 35.13,	p < .001
4 year scholarship recipients	>	2 year former recipients	F(1, 1365) = 13.68,	p < .001
1 year former recipients	>	2 year former recipients	F(1, 1365) = 7.47,	p = .006

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Reading Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year former recipients	F(1, 3327) = 12.44,	p < .001
Public non-applicants	>	3 year former recipients	F(1, 3327) = 17.36,	p < .001
Public scholarship winner non-users	>	3 year former recipients	F(1, 3327) = 12.86,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Reading Achievement	Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	1 year former recipients	F(1, 3327) = 10.71,	p = .001
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 9.77,	p = .002
Public applicant non-recipients	>	3 year former recipients	F(1, 3327 = 17.08,	p < .001
Public non-applicants	>	1 year former recipients	F(1, 3327) = 10.01,	p = .002
Public non-applicants	>	2 year former recipients	F(1, 3327) = 8.88,	p = .003
Public non-applicants	>	3 year former recipients	F(1, 3327 = 15.97,	p < .001
Public scholarship winner non-users	>	1 year former recipients	F(1, 3327) = 12.38,	p < .001
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 11.24,	p < .001
Public scholarship winner non-users	>	3 year former recipients	F(1, 3327 = 17.90,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Reading Achievement	Late 1st to Late 2nd	

Groups			F-statistic	p-value
3 year former recipients	>	Public applicant non-recipients	F(1, 3327) = 6.93,	p = .009
3 year former recipients	>	Public non-applicants	F(1, 3327) = 7.51,	p = .006
3 year former recipients	>	Public scholarship winner non-users	F(1, 3327) = 8.58,	p = .003

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Reading Achievement	Late 2nd	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 12.41,	p < .001
Public non-applicants	>	2 year former recipients	F(1, 3327) = 10.33,	p = .001
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 12.86,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Reading Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 20.27,	p < .001
Public non-applicants	>	2 year former recipients	F(1, 3327) = 19.99,	p < .001
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 15.18,	p < .001

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Language Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year former recipients	F(1, 3327) = 8.24, p = .004	
Public non-applicants	>	3 year former recipients	F(1, 3327) = 9.32, p = .002	
Public scholarship winner non-users	>	3 year former recipients	F(1, 3327) = 10.70, p = .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Language Achievement	Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year former recipients	F(1, 3327) = 20.73, p < .001	
Public non-applicants	>	3 year former recipients	F(1, 3327) = 22.48, p < .001	
Public scholarship winner non-users	>	3 year former recipients	F(1, 3327) = 12.84, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Language Achievement	Late 1st to Late 2nd	

Groups			F-statistic	p-value
3 year former recipients	>	Public applicant non-recipients	F(1, 3327) = 7.85, p = .005	
3 year former recipients	>	Public non-applicants	F(1, 3327) = 8.04, p = .005	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Language Achievement	Late 2nd to Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 12.65 p <.001	
Public non-applicants	>	2 year former recipients	F(1, 3327) = 13.29, p <.001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Language Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 39.78, p < .001	
Public non-applicants	>	2 year former recipients	F(1, 3327) = 43.34, p < .001	
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 24.12, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Math Achievement	Early 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year former recipients	F(1, 3327) = 8.69, p = . 003	
Public non-applicants	>	3 year former recipients	F(1, 3327) = 12.43, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Math Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public non-applicants	>	3 year former recipients	F(1, 3327) = 9.61, p = .002	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Math Achievement	Late 1st	

Groups			F-statistic	p-value
Public non-applicants	>	1 year former recipients	F(1, 3327) = 7.06, p = .008	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Math Achievement	Late 2nd	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 16.01, p < .001	
Public non-applicants	>	2 year former recipients	F(1, 3327) = 25.76, p < .001	
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 10.10, p = .002	
Public applicant non-recipients	>	1 year former recipients	F(1, 3327) = 6.63, p = .010	
Public non-applicants	>	1 year former recipients	F(1, 3327) = 11.14, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Math Achievement	Late 3rd Grade (Spring 2001)	

Groups			F-statistic	p-value
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 28.23, p < .001	
Public non-applicants	>	2 year former recipients	F(1, 3327) = 38.57, p < .001	
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 23.20, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Total Achievement	Early 1st to Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	3 year former recipients	F(1, 3327) = 8.08, p = .005	
Public non-applicants	>	3 year former recipients	F(1, 3327) = 10.87, p = .001	
Public scholarship winner non-users	>	3 year former recipients	F(1, 3327) = 10.94, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Total Achievement	Late 1st	

Groups			F-statistic	p-value
Public applicant non-recipients	>	1 year former recipients	F(1, 3327) = 8.02, p = .005	
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 9.21, p = .002	
Public applicant non-recipients	>	3 year former recipients	F(1, 3327 = 26.24, p < .001	
Public non-applicants	>	1 year former recipients	F(1, 3327) = 7.44, p = .006	
Public non-applicants	>	2 year former recipients	F(1, 3327) = 7.70, p = .006	
Public non-applicants	>	3 year former recipients	F(1, 3327 = 21.17, p < .001	
Public scholarship winner non-users	>	1 year former recipients	F(1, 3327) = 9.50, p = .002	
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 11.49, p < .001	
Public scholarship winner non-users	>	3 year former recipients	F(1, 3327 = 30.10, p < .001	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Total Achievement	Late 1st to Late 2nd	

Groups			F-statistic	p-value
3 year former recipients	>	Public applicant non-recipients	F(1, 3327) = 12.94, p < .001	
3 year former recipients	>	Public non-applicants	F(1, 3327) = 12.41, p < .001	
3 year former recipients	>	Public scholarship winner non-users	F(1, 3327) = 9.67, p = .002	

Measure:	Testing Episode / Interval	Former scholarship recipients v. Public school groups
Total Achievement	Late 2nd	

Groups			F-statistic	p-value
Public applicant non-recipients	>	1 year former recipients	F(1, 3327) = 9.13, p = .003	
Public non-applicants	>	1 year former recipients	F(1, 3327) = 11.42, p < .001	

Public scholarship winner non-users	>	1 year former recipients	F(1, 3327) = 8.51, p = .004
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 15.71, p < .001
Public non-applicants	>	2 year former recipients	F(1, 3327) = 20.02, p < .001
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 12.93, p < .001
Measure:	**Testing Episode / Interval**		Former scholarship recipients v. Public school groups
Total Achievement	Late 3rd Grade (Spring 2001)		
Groups			**F-statistic** **p-value**
Public applicant non-recipients	>	2 year former recipients	F(1, 3327) = 55.12, p < .001
Public non-applicants	>	2 year former recipients	F(1, 3327) = 62.77, p < .001
Public scholarship winner non-users	>	2 year former recipients	F(1, 3327) = 39.27, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Language	End 1st		
Groups			**F-statistic** **p-value**
Scholarship recipients	>	Former scholarship recipients	F(1, 1365) = 14.51, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Language	End 3rd		
Groups			**F-statistic** **p-value**
Scholarship recipients	>	Former scholarship recipients	F(1, 1365) = 50.55, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Language	End 1st to End 2nd		
Groups			**F-statistic** **p-value**
Former scholarship recipients	>	Scholarship recipients	F(1, 1365) = 5.72, p = .017
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Language	End 1st to End 3rd		
Groups			**F-statistic** **p-value**
Former scholarship recipients	>	Scholarship recipients	F(1, 1365) = 5.72, p = .017
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Mathematics	Early 1st		
Groups			**F-statistic** **p-value**
Scholarship recipients	>	Former scholarship recipients	F(1, 1365) = 13.95, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Mathematics	End 1st		
Groups			**F-statistic** **p-value**
Scholarship recipients	>	Former scholarship recipients	F(1, 1365) = 5.56, p = .019
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Mathematics	End 1st to End 3rd		
Groups			**F-statistic** **p-value**
Former scholarship recipients	>	Scholarship recipients	F(1, 1365) = 10.33, p = .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Total	End 1st		
Groups			**F-statistic** **p-value**
Scholarship recipients	>	Former scholarship recipients	F(1, 1188) = 10.60, p = .001

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Total	End 3rd	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Scholarship recipients	>	Former scholarship recipients	F(1, 1188) = 46.26, p < .001	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Total	End 1st to End 3rd	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Former scholarship recipients	>	Scholarship recipients	F(1, 1188) = 15.79, p < .001	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Reading	End 1st	Current Scholarship Recipient & Former Recipients at the time vs. Public School Students

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 19.64, p < .001	
Public school students	>	Scholarship recipients	F(1, 3327) = 25.07, p < .001	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Reading	End 2nd	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 6.00, p = .014	
Public school students	>	Scholarship recipients	F(1, 3327) = 4.96, p = .026	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Reading	End 3rd	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 4.75, p = .029	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Language	1st	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 4.30, p = .038	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Language	End 3rd	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 20.59, p < .001	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Mathematics	Early 1st	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 8.61, p < .003	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Mathematics	End 1st	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 7.05, p = .008	
Public school students	>	Scholarship recipients	F(1, 3327) = 7.03, p = .008	

Measure:	Testing Episode / Interval	Orthogonal Comparisons
Mathematics	End 2nd	Current Scholarship Recipient vs. Former Recipients at the time

Groups			F-statistic	p-value
Public school students	>	Former scholarship recipients	F(1, 3327) = 6.43, p = .011	

Public school students	>	Scholarship recipients	F(1, 3327) = 7.76, p = .005
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Mathematics	End 3rd		
Groups			**F-statistic** **p-value**
Public school students	>	Former scholarship recipients	F(1, 3327) = 10.72, p = .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Total	End 1st		
Groups			**F-statistic** **p-value**
Public school students	>	Former scholarship recipients	F(1, 3327) = 29.52, p < .001
Public school students	>	Scholarship recipients	F(1, 3327) = 20.74, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Total	End 2nd		
Groups			**F-statistic** **p-value**
Public school students	>	Former scholarship recipients	F(1, 3327) = 9.84, p = .002
Public school students	>	Scholarship recipients	F(1, 3327) = 10.41, p = .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons Current Scholarship Recipient vs. Former Recipients at the time
Total	End 3rd		
Groups			**F-statistic** **p-value**
Public school students	>	Former scholarship recipients	F(1, 3327) = 21.08, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons 3-year Former Recipients vs. Other Former Recipients & Public School Students
Reading	End 1st		
Groups			**F-statistic** **p-value**
Public school students	>	3-year Former scholarship recipients	F(1, 3327) = 19.64, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons 3-year Former Recipients vs. Other Former Recipients & Public School Students
Reading	End 3rd		
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 10.51, p = .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons 3-year Former Recipients vs. Other Former Recipients & Public School Students
Reading	End 1st to End 3rd		
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 16.55, p < .001
3-year Former scholarship recipients	>	Public school students	F(1, 3327) = 6.38, p = .012
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons 3-year Former Recipients vs. Other Former Recipients & Public School Students
Language	End 1st		
Groups			**F-statistic** **p-value**
Public school students	>	3-year Former scholarship recipients	F(1, 3327) = 20.93, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons 3-year Former Recipients vs. Other Former Recipients & Public School Students
Language	End 3rd		
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 11.09, p = .001

Measure:	Testing Episode / Interval		Orthogonal Comparisons
Language	End 1st to End 3rd		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 15..95, p < .001
3-year Former scholarship recipients	>	Public school students	F(1, 3327) = 9.94, p = .002
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons
Mathematics	End 1st		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
Public school students	>	3-year Former scholarship recipients	F(1, 3327) = 7.05, p = .008
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons
Mathematics	End 3rd		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 15.07, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons
Mathematics	End 1st to End 3rd		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 9.94, p = .002
3-year Former scholarship recipients	>	Public school students	F(1, 3327) = 7.24, p = .007
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons
Total	End 1st		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
Public school students	>	3-year Former scholarship recipients	F(1, 3327) = 29.41, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons
Total	End 3rd		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 23.15, p < .001
Measure:	**Testing Episode / Interval**		Orthogonal Comparisons
Total	End 1st to End 3rd		3-year Former Recipients vs. Other Former Recipients & Public School Students
Groups			**F-statistic** **p-value**
3-year Former scholarship recipients	>	Other Former scholarship recipients	F(1, 3327) = 19.36, p < .001
3-year Former scholarship recipients	>	Public school students	F(1, 3327) = 21.19, p < .001

Appendix C

Regression toward the Mean: An Alternative Explanation

Regression toward the mean is a statistical artifact that can obfuscate the findings of analyses designed to examine change across time. A simple definition of regression toward the mean is that, on average, any given individual is going to be closer to the mean at *time two* than he or she was at *time one*. For instance, when achievement measures are obtained at two separate points in time, students with the highest scores at the end of second grade are, on average, likely to display lower scores at the end of third grade (i.e., closer to the mean of the population to which they belong). In contrast, students with the lowest scores at the end of second grade are, on average, likely to display higher scores at the end of third grade (i.e., closer to the mean of the population to which they belong).

Regression toward the mean tends to occur because a point estimate (a single measure) of an individual's ability is probabilistically less representative of an individual's true ability than is the mean obtained across multiple measures. When multiple measures of ability (or any other construct of interest) are obtained from an individual, there always will be a certain amount of variance among the scores, such that some scores will be higher than others. For any single score from an individual, 50% of the time it will fall above the individual's true mean, and 50% of the time it will fall below the individual's true mean. Therefore, when an individual's ability is measured multiple times, the scores will form a distribution with a mean that falls between the individual's high and low scores. The mean score is a better estimate of this person's true ability than is any single score in the distribution (i.e., an individual's true ability is best represented by the mean obtained across multiple measures). Consequently, if a score obtained at time one is greater than an individual's true mean-level of ability, the chances are that this individual's score will be lower at time two, yielding an average for the two obtained scores that is closer to the individual's true average ability level. In other words, multiple scores will mathematically converge on the individuals' average across time – a statistical artifact know as regression toward the mean.

Regression toward the mean in longitudinal studies can be problematic when differences in initial levels of performance vary systematically across the levels of a particular variable of interest (e.g., participation in the CSTP). In terms of the present study, it may appear as though change has occurred across time as function of participation in the program when, in actuality, the change may be attributable to regression effects. Although recent researchers have argued that regression effects are not as ubiquitous as previously believed,[79] the effect can be problematic with intact groups that have not been randomly assigned to control for differences in initial performance.

According to Rogosa (1995), researchers conducting longitudinal investigations must be wary of regression toward the mean effects when the correlation between change across time and initial status (e.g., achievement at the beginning of first grade) is *negative* and large in magnitude.[80] Therefore, when greater (less) change across time is strongly related to lower (higher) initial achievement, *and* initially high and low performing students are not randomly distributed among the treatment groups (e.g., scholarship status groups), then regression toward the mean cannot be ruled out as a plausible alternative explanation for differential change in achievement scores across time between one or more of the treatment groups.

In the present study, the results of the achievement analyses indicate that the initially high performing students at the beginning of first grade (autumn, 1998) are predominantly in the three-year and four-year scholarship recipient groups, whereas the initially low performing students are disproportionately represented in one or more of the public school groups. Thus, initial achievement levels are confounded with the scholarship recipient groups. In other words, initial achievement varies systematically, as opposed to randomly, across the different groups in the analyses. Even more problematic, however, is the finding that the observed initial achievement scores and the corresponding measures of overall achievement change from first

[79] See Rogosa, D. R. (1995). Myths and methods: Myths about longitudinal research plus supplemental questions. In The analysis of change, J. M. Gottman, Ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 3-65.

[80] Rogosa (1995, see reference in the footnote above) emphasizes and mathematically demonstrates that the correlation between change and initial status in a longitudinal study can be negative, positive, or zero, whereas many researchers believe that the correlation is always negative, regardless of whether regression toward the mean is at play. Therefore, a strong negative correlation between initial status and change can serve as a useful indicator of the presence of regression toward the mean. Rogosa, however, stresses that a strong negative correlation between change and initial status is a necessary but NOT a sufficient condition for regression effects. This is because the correlation tends to be negatively biased, such that negative correlations between observed change and observed initial status can be found when the "true-score" correlation is zero or positive.

grade to third grade are strongly, and negatively, correlated on all of the achievement measures (see Table 62 below). Therefore, initially high achieving students tend to display less achievement change across time than do initially low achieving students.

Because the initially high performing students are disproportionately represented among three-year and four-year scholarship recipients, it is not possible to determine whether these students demonstrate less change across time because of factors specific to the CSTP, or simply because of the effect of regression toward the mean. A similar issue arises when attempting to interpret the greater amount of achievement change displayed by the initially low-performing public school comparison groups. Consequently, some caution should be exercised when interpreting any of the findings that are indicative of differential achievement change across time between two or more of the student groups.

Table 62. Correlations between Initial Achievement Scores (Early First Grade, 1998) and Achievement Change from the Beginning of First Grade to the end of Third Grade

	Reading	Language	Mathematics	Total
Correlation between early 1st Grade Achievement & Overall Achievement Change from 1st to 3rd Grade	**- .43**	**- .56**	**- .48**	**- .35**
p-value	< .0001	< .0001	< .0001	< .0001

N = 1813

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Evaluation of the Cleveland Scholarship and Tutoring Program 1998-2001 Executive Summary, Summary Report & Technical Reports	
Author(s): Kim K. Metcalf, Stephen D. West, Natalie A. Legan, Kelli M. Paul, William J. Boone	
Corporate Source: Indiana Center for Evaluation @ Indiana University	Publication Date: 10/2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒ Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Level 2A ☐ Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Level 2B ☐ Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: [signature]	Printed Name/Position/Title: Natalie A. Legan / Research Associate
Organization/Address: 509 E. 3rd Street, Bloomington, IN. 47401	Telephone: 800 511-6575 FAX: 812856-5890
	E-Mail Address: nalegan@indiana.edu Date: 6/25/03

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:	Document can also be accessed @ www.indiana.edu/~ivice via the "reports" Link.
Address:	you do not need to cite this - only do so if you all desire
Price:	

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)